

UNIVERSITE DE PARIS I PANTHEON-SORBONNE
U.F.R. DE SCIENCES ECONOMIQUES

Année 2008

Numéro attribué par la bibliothèque

2008PA10022

THESE

pour obtenir le grade de

Docteur en Sciences Economiques de l'Université de Paris I

Présentée et soutenue publiquement par

Patrick-Serge Kouontchou Kepawou

le 20 juin 2008

Titre :

DE L'INFORMATION EN HAUTE FREQUENCE : QUATRE ESSAIS
EMPIRIQUES SUR LES MESURES DE RISQUES ET
L'EVALUATION DES ACTIFS FINANCIERS

Sous la direction de Monsieur le Professeur Thierry Chauveau

(Université de Paris I - CES/CNRS)

Jury de thèse

M. Thierry Chauveau (Directeur de thèse),	Professeur à l'Université de Paris I
Mme Hélène Raymond-Feingold (Rapporteur),	Professeur à l'Université de Paris X
M. Christophe Hurlin (Rapporteur),	Professeur à l'Université d'Orléans
M. Gunther Capelle-Blancard (Président),	Professeur à l'Université de Paris I
M. Bertrand Maillet (Suffragant),	Maître de Conférences à l'Université de Paris I

L'université de Paris I Panthéon-Sorbonne n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

*A mon père Albert, à ma mère Pauline et à ma belle-mère Jacqueline, qui n'ont jamais
ménagé leurs efforts pour veiller à mon épanouissement, je dédie cette thèse.*

Remerciements

Je tiens tout d'abord à remercier Monsieur le Professeur Thierry Chauveau pour m'avoir donné l'occasion de poursuivre mes études de doctorat dans le cadre prestigieux de l'université de Paris I. Je lui suis très reconnaissant de la confiance qu'il a su me témoigner au cours de ces années, tout en me faisant partager son expérience, sa curiosité et son enthousiasme toujours renouvelé.

Je voudrais également témoigner ici ma reconnaissance à Bertrand Maillet qui a encadré cette thèse. Sans lui, ce travail n'aurait jamais pu voir le jour. Au delà de son accueil et des différentes opportunités qu'il m'a offertes, ses conseils et sa rigueur au travail m'ont permis de me surpasser et de trouver à chaque fois l'énergie suffisante pour résoudre les problèmes qui se posaient. Qu'il trouve dans ce travail le fruit de ses sacrifices et de sa confiance. Nous avons travaillé ensemble sur des sujets qui sont repris dans les différents chapitres de cette thèse. J'ose espérer que notre collaboration, qui va bien au-delà de cette thèse, continuera dans les années à venir.

Ma reconnaissance va, bien entendu, à Monsieur le Professeur Christian Bordes qui m'a accepté dans son cycle de DEA Monnaie-Banque-Finance après ma formation initiale d'Ingénieur Statisticien. Cette opportunité m'a permis d'acquérir de nouvelles connaissances en finance et de faire mes premiers pas dans la recherche au travers d'une Bourse de mérite de la Coopération Française.

Je veux remercier vivement Madame le Professeur Hélène Raymond-Feingold (Université

de Paris X) et Monsieur le Professeur Christophe Hurlin (Université d'Orléans) qui ont accepté de me faire l'honneur d'être rapporteurs de cette thèse; ainsi que Monsieur le Professeur Gunther Capelle-Blancard (Université de Paris I) pour avoir accepté de participer au jury.

Ma gratitude va ensuite à Monsieur le Professeur Michel Verleysen et toute son équipe de recherche, qui m'ont chaleureusement accueilli plusieurs fois (été 2005 et 2006) dans leur laboratoire de Micro-Electronique de l'Université de Louvain-La-Neuve (Belgique). Ils m'ont ainsi initié aux techniques d'extraction des signaux par l'Analyse en Composantes Indépendantes, ce qui m'a permis d'avancer de façon significative dans le chapitre consacré aux mesures de risques en haute fréquence.

Je voudrais aussi remercier vivement toute l'équipe AAAdvisors (ABN AMRO), plus particulièrement Monsieur Stéphane Corsaletti, pour m'avoir permis de travailler en lien direct avec son service de recherche quantitative depuis le début de cette thèse. J'y ai beaucoup appris sur la pratique quotidienne de la finance. Les échanges professionnels avec le service de gestion ont toujours été très enrichissants et utiles pour l'accomplissement de cette thèse.

Je pense également à Thierry Michel (avec qui nous avons collaboré pour le travail sur la déformation du temps), Emmanuel Jurczenko, Christophe Boucher et Jean-Philippe Médecin, pour m'avoir fait partager leurs expériences dans le domaine de la recherche quantitative financière. Leurs remarques, suggestions et collaborations ont été toujours très enrichissantes. Je voudrais qu'ils sachent aussi que leurs conseils et encouragements m'ont été d'une grande utilité pour la réalisation de ce travail.

Je tiens à remercier également Paul Merlin, Benjamin Hamidi, Guillaume Bagnarosa, Ghislain Yanou, Caroline Barrault et Rachid Bokreta pour leur assistance constante et l'ambiance studieuse durant ces dernières années. J'aimerais qu'ils sachent que ces années

passées ensemble ont été riches d'enseignements ; que ce soit lors de nos rencontres hebdomadaires de recherche à Vincennes, ou lors des discussions en tête-à-tête sur des sujets parfois éloignés de mon domaine de recherche.

J'adresse aussi mes sincères remerciements aux participants des conférences et séminaires auxquels j'ai participé : GdR Economie Monétaire et Financière (Strasbourg, Juin 2005), Conférence internationale de l'AFFI (Paris, Juin 2005 ; Bordeaux, Juin 2007), Journée d'Econométrie Financière Avancée (Paris, Décembre 2005, Décembre 2006 et Décembre 2007), Journées de Micro-économie Appliquée (JMA06 - Nantes, Juin 2006), *European Symposium on Artificial Neural Networks* (ESANN07 - Bruges, Avril 2007). Qu'ils trouvent ici une trace des remarques pertinentes formulées lors de mes présentations.

Je témoigne toute ma gratitude à Dieudonné Kinkielele et à Robert Ngonthé, directeurs des études à l'Institut Sous-régional de Statistiques et d'Economie Appliquée (ISSEA) de Yaoundé au Cameroun, pour l'enthousiasme et la ferveur au travail qu'ils mettent au service des étudiants de cet institut. Les multiples séjours qu'ils m'ont permis de faire dans leur école, pour participer à des jurys de soutenance des mémoires des étudiants en fin de cycle, ont toujours été très valorisants et pleins d'enseignements.

Mes remerciements vont aussi évidemment à mes proches : mes frères et soeurs, Marie-claire Mendom, Pascaline Simeu, Suzanne Kembou, famille Fankem, Bruno Tchouanfang et Romuald Kenmoe. A mes amis, Octavie Mbiakoup, Josiane Guedem, Augustave Ngouné, Armel Foléack, Charlie Tchinda et particulièrement en ce moment, je pense à Pelagie Kamta qui a su, durant ces années de recherche, être patiente et compréhensive. Qu'ils sachent tous que ce travail est arrivé à son terme grâce à leur indéfectible soutien.

Ce travail a bénéficié, dans sa version finale, des remarques et suggestions de plusieurs lecteurs. Les éventuelles erreurs, omissions ou coquilles qui resteront ne pourront leurs être imputées et seront donc entièrement à la charge de l'auteur.

Table des matières

Remerciements	iii
Table des figures	ix
Liste des tableaux	xii
Introduction Générale	1
Résumé du chapitre sur « Roses des vents », « éventails » et « explosions d'étoiles » sur le marché français : caractérisations, mesures et applications	5
Résumé du chapitre sur “ <i>High-frequency Market Risk Scenarii by Independent Component Analysis: Method and Application</i> ”	15
Résumé du chapitre sur “ <i>A Robust Conditional Multi-moment Asset Pricing Model with High-frequency Data</i> ”	26
Résumé du chapitre sur “ <i>Recovering an Intrinsic Well-behaved Business Time from Market Prices</i> ”	35
1 « Roses des vents », « éventails » et « explosions d'étoiles » sur le marché français : caractérisations, mesures et applications	45
1.1 Introduction	45
1.2 Définition, caractérisation et mesures d'une rose des vents	51

1.2.1	Définition et caractérisation du phénomène	51
1.2.2	De la mesure du phénomène à la construction des indices d'intensité du phénomène de rose des vents (<i>Compass Rose Intensity Indexes</i>)	57
1.3	Applications et caractérisations empiriques du phénomène	63
1.3.1	Conditions d'existence sur données simulées	63
1.3.2	Conditions d'existence sur données de marché	80
1.4	Conclusion	85
2	High-frequency Market Risk <i>Scenarii</i> by Independent Component Analysis: Method and Application	91
2.1	Introduction	91
2.2	Independent Component Analysis	93
2.2.1	Definition	93
2.2.2	Estimation of the ICA model	96
2.2.3	FastICA Algorithm	99
2.2.4	An Example of ICA Estimation	100
2.2.5	Reasons for Exploring ICA with Financial Time-series	101
2.3	Value-at-Risk Calculation	102
2.3.1	General Framework of VaR Estimation	102
2.3.2	Computing VaR with ICA	104
2.3.3	Backtesting VaR Models	106
2.4	Experiments and Results	110
2.5	Conclusion	120
3	A Robust Conditional Multi-moment Asset Pricing Model with High-frequency Data	127
3.1	Introduction	127
3.2	Factor Pricing Models and Realized Co-variations	130
3.2.1	Factor Pricing Models	130
3.2.2	Realized Loadings and Co-variations	132

3.2.3	Realized Betas on the French Stock Market	134
3.3	Conditional Asset Pricing Model	139
3.3.1	Realized Conditional Asset Pricing Model with Higher Moments .	139
3.3.2	Regressions Tests of the CAPM	142
3.3.3	Estimation Procedure of the Model	146
3.4	Results and Robustness Tests	149
3.4.1	Fama-French Realized Portfolio Returns and Factor Loadings . .	149
3.4.2	High-frequency Factor Loadings and Asset Pricing	152
3.5	Conclusion	164
4	Recovering an Intrinsic Well-behaved Business Time from Market Prices	170
4.1	Introduction	170
4.2	Recovering an Intrinsic Time	172
4.3	Characteristics of the Intrinsic Business Times	180
4.4	A Common Business Time for the Stock Market	187
4.5	Conclusion	191
	Conclusion Générale	193
	Annexe : On the Impact of the Compass Rose Phenomenon on GARCH	
	Estimation	201
	Références bibliographiques	211

Table des figures

1.1	Diagramme de phases des rentabilités	47
1.2	Diagramme de phases des rentabilités du titre Peugeot	52
1.3	Diagramme de phases des rentabilités du titre Peugeot avec modification de l'échelle	53
1.4	Phénomène de rose des vents pour les rentabilités au carré (illustration d'un éventail)	54
1.5	Phénomène de rose des vents pour des triplets de rentabilités successives (explosion d'étoiles)	55
1.6	Histogramme des valeurs normalisées du θ	60
1.7	Phénomène de rose des vents pour différentes tailles d'échelon de cotation (sans corrélation)	65
1.8	Phénomène de rose des vents pour différentes tailles d'échelon de cotation (avec une structure de corrélation imposée)	68
1.9	Diagramme de phases des rentabilités pour différentes tailles de l'échantillon	69
1.10	Indices d'intensité de rose des vents pour différentes tailles de l'échantillon	70
1.11	Diagramme de phases des rentabilités pour différentes tailles de l'échelon et de volatilité de la série - modèle <i>AR-GARCH</i> -	73
1.12	Eventail selon l'échelon de cotation - modèle <i>AR-GARCH</i> -	74
1.13	Phénomène d'explosion d'étoiles - modèle <i>AR-GARCH</i> -	76
1.14	Diagramme de phases des rentabilités pour différentes tailles de l'échantillon - modèle <i>AR-Student-GARCH</i> -	77

1.15	Diagramme de phases des rentabilités pour différentes tailles de l'échantillon - modèle <i>AR-Skew-Normal-GARCH</i> -	78
1.16	Diagramme de phases des rentabilités pour différentes tailles de l'échantillon - modèle <i>AR-Skew-Student-GARCH</i> -	79
1.17	Rose des vents, éventail et explosion d'étoiles des différents titres cotés sur le marché français des actions	83
1.18	Moyennes des volatilités réalisées, des biais d'auto-corrélation et des me- sures de la rose des vents en fonction de la fréquence d'échantillonnage .	87
2.1	Schematic Representation of ICA	94
2.2	Illustration of Independent Component Analysis Extraction Technique .	101
2.3	Independent and Principal Components with corresponding Densities for 10 and 15-minute quotes	113
2.4	A VaR Forecasts Comparison at 95% confidence threshold	114
2.5	A VaR Forecasts Comparison at 97.5% confidence threshold	115
2.6	A VaR Forecasts Comparison at 99% confidence threshold	116
2.7	A VaR Forecasts Comparison at 99.5% confidence threshold	117
3.1	Average Adjusted Factor Loadings for smaller size and style Portfolio . .	138
3.2	Realized Adjusted Factor Loadings of Large Size and Higher Style Portfolio	139
4.1	Convergence of the Anderson-Darling Objective Function	178
4.2	Convergence of the Solutions towards a Single Genome	179
4.3	Calendar (top) and Business (bottom) Returns for <i>Air Liquide</i>	181
4.4	Log-densities for <i>Air Liquide</i> calendar and business returns	182
4.5	Moment-based Normality Diagnoses	183
4.6	P-P Plot Normality Diagnosis for Air Liquide Returns in Calendar Time (black) and Business Time (red)	183
4.7	Number of Daily (top) and Weekly (bottom) Sampled Observations for <i>Air Liquide</i>	184

4.8	Distribution of the Log-durations of the 43 Business-time Clocks	184
4.9	Envelope of Business Clocks over the Sample of 43 Stocks	186
4.10	Correlations in Calendar Time (bottom) and Business Time (top)	186
4.11	P-P Plot of Standard Returns in Calendar (Black) and Business (Red) Times	187
4.12	Log-Densities of Standard Returns in Calendar and Business Times	188
4.13	Pair-wise Correlations in Business Time <i>versus</i> Calendar Time	189
4.14	Clustering of Stocks in Calendar and Business Times	190
4.15	Out-of-sample Absolute Returns of the Minimum Variance Portfolios	190
4.16	Distribution Function of Out-of-Sample Returns of the Minimum Variance Portfolios	191

Liste des tableaux

1.1	Caractéristiques des prix des titres considérés	80
1.2	Tableau descriptif des rendements des titres de l'échantillon	81
1.3	Auto-corrélations d'ordre 1 des rendements (et transformés) et statistiques descriptives sur les durations	82
1.4	Intensité du phénomène de rose des vents pour les titres considérés . . .	84
2.1	Summary Statistics of Independent Factors from ICA	111
2.2	Summary Statistics of Factors from PCA	112
2.3	Value-at-Risk Measures of Portfolio Returns	118
2.4	Conditional Value-at-Risk (CVaR) of an Equally Weighted Portfolio for different VaR-based Measures	119
2.5	Observed Frequency of Exceptions	122
2.6	Unconditional Coverage Test Results	123
2.7	Independent Test Results	124
2.8	Conditional Coverage Test Results	125
2.9	Dynamic Quantile Test Results	126
3.1	Summary Statistics for IntraDaily Returns	135
3.2	Mean, Realized Volatility for Monthly Returns	136
3.3	Realized Factor Loadings	150
3.4	Average R^2 for Alternatives Models with or without Skewness Factor . .	151
3.5	Factor Loading Forecasts	154
3.6	Average Pricing Errors and Significance Tests	158

3.7	Average of $Pseudo-R^2$ and Significance Tests	159
3.8	Pseudo R^2 and Significance Test of Portfolios	160
3.9	Mean-variance Efficient Portfolios	162
3.10	Willingness-to-Pay for Different Strategies	163
4.1	A Sample Genome (for the stock $TF1$)	173
4.2	Mutation of a Genome	175
4.3	One Point Cross-over	176

Introduction Générale

Depuis une vingtaine d'années, les marchés financiers ont évolué de manière spectaculaire, stimulés par la déréglementation, le découplage et la désintermédiation du système. La multiplication des produits financiers et l'internationalisation des marchés sont venues répondre aux besoins d'allocation *optimale* de l'épargne au niveau mondial et ont permis le développement des opérations d'arbitrage, de couverture et de spéculation. Cet accroissement de l'offre de produits financiers a fortement modifié les comportements des agents économiques. En effet, jusqu'au début des années 70, une majorité d'épargnants recherchait des placements liquides sans risques et contre partie, faiblement rémunérés. Après les chocs pétroliers, les mouvements inflationnistes qui ont suivi et la multiplication des produits financiers, ont progressivement amené ceux-ci à changer leur profil d'investissement pour rechercher des produits plus risqués certes mais aussi plus rémunérateurs. Parallèlement à cette modification du comportement des agents, les marchés financiers ont opéré des modifications structurelles importantes pour faire face à la demande. Pour la plupart, ils sont passés du mode de cotation en *fixing*, où les transactions et les collectes des ordres étaient effectuées sur des intervalles de temps fixés par avance, à la cotation en continue, où les échanges sont enregistrés de manière quasi-continue (chaque seconde pour les titres les plus liquides).

Ce changement du mode de cotation a été rendu possible grâce au développement de l'informatique et l'arrivée de la technologie électronique qui ont réduit considérablement les coûts de collecte et de traitement des informations. Elle a aussi permis de transmettre simultanément les informations en temps réel aux participants des marchés. Ces change-

ments structurels dans les moyens de collecte, de traitement et de transmission de l'information ont eu des implications importantes pour la disponibilité et l'interprétation des données de haute fréquence. En effet, la cotation en temps continu permet de disposer pratiquement de tous les ensembles de données sur l'activité du marché (prix des actifs, volumes échangés, nombres d'acteurs, nombre de vendeurs, fréquence d'échange, ...). Ces informations plus proches des processus générateurs des prix des actifs sont sans doute plus appropriées pour les tests empiriques des modèles théoriques de la littérature financière. Ceci d'autant plus que la plupart des études empiriques n'arrivent pas à valider certains résultats théoriques à partir des données dites de basse fréquence (journalières, hebdomadaires ou mensuelles ; Cf. Fama et French, 1996). L'arrivée des données de haute fréquence peut donc combler ce déficit et permettre d'enrichir la littérature sur la modélisation des dépendances entre les actifs de façon robuste. Toutefois, ces données sont-elles le reflet réel des dépendances entre les sources potentielles de risque ? Ont-elles les mêmes caractéristiques que les données de basse fréquence ? Peuvent-elles être utilisées sur des modèles de dépendances classiques ? Peuvent-elles mieux rendre compte des comportements des agents sur les marchés ?

Un grand nombre d'études - principalement anglo-saxonnes - portant sur ce type de données a été récemment publié. Ces études mettent en avant certaines caractéristiques importantes des prix des actifs financiers, jusqu'alors pratiquement indécélables avec des données de basse fréquence. C'est pourquoi la constitution et l'accès à des bases de données de haute fréquence constituent un enjeu important pour analyser et mieux comprendre les marchés financiers (Wood *et alii*, 1985). C'est en effet une possibilité supplémentaire d'approfondir la connaissance des éléments fondamentaux qui dirigent les prix des actifs. En se rapprochant du « vrai » processus gouvernant les prix, les données de haute fréquence permettent de revisiter les modèles traditionnels de la finance et les relations d'équilibre qu'ils sous-tendent. Néanmoins, le traitement de ces données spécifiques ne va pas sans poser un certain nombre de problèmes. Notamment, les bruits associés aux données - liés, entre autres, à des problématiques d'organisation des marchés et à des phénomènes de microstructure - doivent faire l'objet d'un traitement particu-

lier (Feinstone, 1987). De plus, les hypothèses distributionnelles classiques (normalité des distributions, linéarités des relations,...), plutôt vérifiées en données de basse fréquence, doivent être modifiées pour tenir compte des particularités de la haute fréquence. L'objectif de cette thèse est de mieux exploiter le contenu informationnel de ces données de haute fréquence, tout en évaluant et corrigeant leur impact sur les modèles classiques proposés dans la littérature financière. Ces corrections porteront principalement sur le choix des fréquences d'échantillonnage *optimales* et sur la prise en compte dans les processus d'estimation des biais microstructurels. Ce choix portera aussi sur l'utilisation de modèles appropriés (intégrant les moments d'ordre supérieur et leur non-linéarité) et des méthodes d'estimation robustes (algorithmes génétiques, équations simultanées, corrections des biais d'auto-corrélation,...). Ainsi, les quatre chapitres de cette thèse ont pour objectif de traiter ces différentes problématiques.

Nous commençons par l'exploration d'un phénomène de microstructure dénommé « rose des vents » dans la littérature, caractérisant les échanges à très haute fréquence et dû, en particulier, à l'existence de pas d'échelon de cotation (*tick* en anglais) sur les cours des titres. Cette question a fait l'objet de plusieurs études (*Cf.* Crack et Ledoit, 1996 ; Krämer et Runde, 1997 et Gleason *et alii*, 2000) qui concluent à la présence de biais en haute fréquence. Ces études n'ont toutefois pas caractérisé de manière formelle les conditions d'apparition de ce phénomène. Le premier chapitre de la thèse se propose de quantifier l'existence du phénomène sur le marché français des actions, et de déterminer les seuils à partir desquels les paramètres des modèles financiers classiques sont effectivement biaisés. L'accent sera porté dans ce chapitre sur les conditions d'existence de ce phénomène de microstructure à partir des simulations et des données réelles. Ce sera alors l'occasion de définir trois indicateurs complémentaires de mesure du phénomène de rose des vents.

Nous proposons ensuite une méthode efficiente d'évaluation du risque d'un actif (appelée *VaR* pour *Value-at-Risk* en anglais), par l'exploitation des données de haute

fréquence en termes de précision et de fiabilité de la mesure. Il s'agit ici d'utiliser une technique d'extraction des facteurs communs aux sources potentielles de risque à partir d'une analyse en composantes indépendantes (*Independent Component Analysis* en anglais, Cf. Hyvärinen *et alii*, 2001), mieux adaptée aux caractéristiques intrinsèques des données de haute fréquence. En effet, cette approche a l'avantage, dans un contexte financier, de prendre en compte la non-normalité et la non-linéarité des données (Cf. Back et Weigend, 1997).

Nous poursuivons par une généralisation des modèles traditionnels d'évaluation d'actifs (cadre moyenne-variance) aux moments d'ordre supérieur avec des données de haute fréquence. Cette extension est possible par l'utilisation des mesures dites « réalisées » (Cf. Bollerslev et Zhang, 2003) dans un cadre dynamique avec des méthodes d'estimation robustes (Cf. Zellner, 1962 ; Parks, 1967 et Scholes et Williams, 1977). Cette dernière est particulièrement adaptée aux caractéristiques bien connues des variables financières (queues de distributions épaisses, asymétrie, hétéroscédasticité, ...).

Enfin, dans le quatrième chapitre, nous appliquons une méthode de déformation du temps calendaire en temps de transactions pour lequel les rendements correspondants seront distribués suivant une loi normale. Il s'agit d'exploiter l'idée de Clark (1973) qui stipule que la forme de la distribution des rendements des actifs financiers peut être liée à une superposition de plusieurs processus stochastiques. L'extraction d'une distribution particulière (dans notre cas, la distribution normale) est alors effectuée à partir de l'algorithme génétique. Au final, nous obtenons une fréquence d'échantillonnage en temps dit de « transactions » qui assure, à partir des tests d'hypothèses, la normalité de la distribution des rendements. Une première application vise à la détermination des portefeuilles optimaux à partir de cette nouvelle base de données.

Nous présentons maintenant un résumé plus détaillé des quatre problématiques traitées dans la thèse.

Résumé du chapitre sur « Roses des vents », « éventails » et « explosions d'étoiles » sur le marché français : caractérisations, mesures et applications

Le premier chapitre fait une analyse descriptive d'un phénomène particulier de micro-structure lié à la discrétisation des cours boursiers. Cette discrétisation a pour principale conséquence de faire apparaître, lorsque l'on représente dans un plan l'évolution des rendements entre deux dates successives, une structure complexe de droites émergeant de l'origine. Ce type de configuration, appelée « rose des vents » dans la littérature récente, a été découvert en 1996 par Crack et Ledoit (étude publiée dans la revue *Journal of Finance* sous le titre : « *Robust Structure without Predictability : The Compass Rose Pattern of the Stock Market* »). Cette structure prend la forme de droites espacées, émergeant de l'origine et fuyant vers les principales directions des points cardinaux (d'où le nom de « rose des vents » emprunté aux navigateurs du fait de la similarité avec celle des marins). Ce phénomène a fait depuis l'objet de nombreuses études ; tout d'abord, sur la caractérisation de la structure (Szpiro, 1998 ; Lee *et alii*, 1999 ; Gleason *et alii*, 2000 ; Wang *et alii* 2000 et McKenzie et Frino, 2003), ensuite, sur les biais potentiels pour les paramètres des modèles (Krämer et Runde, 1997 et Amilon et Byström, 2000) et enfin, sur la mise en évidence de la visibilité (Koppl et Nardone, 2001 ; Wang et Wang, 2002 et Mitchell et McKenzie, 2006). Ces études ont déterminé les raisons pour lesquelles cet étrange phénomène apparaissait.

La principale cause de ce type de phénomène, mentionnée par la littérature, est la nécessaire discrétisation des cours de bourse. Toutefois, cette discrétisation ne suffit pas, à elle seule, à expliquer le phénomène de rose des vents sur tous les titres. En effet, alors que Crack et Ledoit (1996) ou bien Szpiro (1998) indiquent que la taille de l'échelon de cotation serait déterminante dans l'apparition d'un tel phénomène sur l'ensemble des titres échangés sur le *New York Stock Exchange* (NYSE). Lee *et alii* (1999) trouvent

qu'aucune structure évidente de rose des vents ne peut être mise en évidence sur l'ensemble des titres du marché à terme américain. Ainsi, la forme d'une rose des vents cache en fait une structure plus complexe nommée nano-structure par Szpiro (1998). Pour ce dernier, seuls les changements discrets de prix sont une condition nécessaire et suffisante à l'apparition d'une rose des vents. Wang *et alii* (2000) montrent que l'échelle d'observation a aussi un impact sur l'apparition du phénomène et Gleason *et alii* (2000) mettent en évidence l'influence de la volatilité sur la visibilité de la rose des vents. Au-delà de l'aspect purement visuel, d'autres études, notamment celles de Gottlieb et Kalay (1985), Krämer et Runde (1997) et Amilon et Byström (2000), abordent les implications de la discrétisation des prix sur les estimations des modèles de la théorie financière. Ainsi, le phénomène de rose des vents peut, à partir d'un seuil (non déterminé explicitement), être la cause de biais sur les estimations des modèles autorégressifs de type *ARCH* (Cf. Amilon et Byström, 2000) et sur les tests standard sur le chaos (Cf. Krämer et Runde, 1997). Ces recherches ont un point commun : trouver le seuil du rapport de la taille de l'échelon de cotation/volatilité des rendements à partir duquel une « rose des vents » apparaît et identifier les conséquences sur les estimations des principaux modèles de la théorie financière (*GARCH*, chaos, tests *BDS*...) et leurs prévisions (Amilon et Byström, 2000). Mais, à notre connaissance, une mesure quantitative du phénomène reste toujours à proposer, malgré les efforts portés par Koppl et Nardone (2001), Wang et Wang (2002) et Mitchell et McKenzie (2006) montrant les effets d'une discrétisation forte des cours dans l'espace de phases. Notre propos dans ce chapitre était de comprendre et de caractériser plus avant le phénomène de « rose des vents » et de mettre en lumière cette singulière et intrigante structure des rendements boursiers, puis de déterminer le seuil des variables susceptibles de mener à l'apparition du phénomène. Pour cela, nous avons d'abord situé ce phénomène dans le contexte des biais microstructurels et défini des indicateurs de mesure.

Ce phénomène de microstructure est étroitement lié aux phénomènes déjà proposés dans la littérature. En effet, l'impact des phénomènes de microstructure des marchés sur le processus de formation des prix des actifs a depuis des décennies fait l'objet de recherches

aussi bien théoriques qu'empiriques. Ces recherches ont montré que les composantes de la fourchette affichée de prix (*Bid-Ask*) dépendaient principalement de trois facteurs : le coût de traitement des ordres (*order processing cost*, Demsetz, 1968) ; le coût d'opportunité d'une position mal diversifiée (*inventory holding cost*, Ho et Stoll, 1983 et Biais, 1993) et le coût de sélection adverse (*adverse information cost*, Glosten et Milgrom, 1985).

Cependant ces composantes théoriques ne permettent pas de comprendre certains phénomènes comme l'existence des seuils ou barrières psychologiques sur les marchés boursiers (De Grauwe et Decupere, 1992 ; Morel et Teïletche, 2001) ou bien encore la concentration des prix autour de certaines valeurs (Gottlieb et Kalay, 1985 ; Harris, 1991 ; Anshuman et Kalay, 1998 ; Huang et Stoll, 1994, 2001 et Capelle-Blancard et Chaudhury, 2007). Plusieurs autres articles - parmi lesquels ceux de Holthausen *et alii* (1987), Aitken et Frino (1996), Chan et Lakonishok (1993, 1995, 1997) et Keim et Madhavan (1995, 1996, 1997) - ont montré que la forme particulière de biais microstructurel, liée à la « danse de la fourchette » des prix, affectait la structure auto-corrélative des rendements des actifs. Ce biais proviendrait de l'asymétrie qui existe entre les blocs vendeurs et les blocs acheteurs. Les blocs acheteurs ayant tendance à faire baisser les prix tandis que les blocs vendeurs les font augmenter, la structure auto-corrélative des rendements serait ainsi modifiée par ces deux effets contraires successifs (Frino *et alii*, 2004).

Le phénomène de « rose des vents » analysé ici est cependant plus complexe que les autres formes de biais. En effet, en étudiant directement les variations de prix, nous prenons en compte l'ensemble des facteurs qui peuvent influencer les prix ; et de ce fait, il intègre aussi les autres formes de biais microstructurels. Ces biais microstructurels liés à la rose des vents étant à l'origine de la modification de la structure auto-corrélative des prix sur certains marchés, une littérature récente, originellement inspirée par les études sur la théorie du chaos, est consacrée à la détermination de la structure intrinsèque des rendements des titres sur les marchés financiers. Il est généralement admis qu'une distribution non uniforme des points représentés indique la présence d'une structure sous-jacente ; de

ce fait, travailler sur des données financières pose le problème de la non continuité des prix de cotation qui peut affecter significativement les paramètres des différentes procédures utilisées pour modéliser les rendements des titres.

En effet, les prix des titres, déterminés par les variations minimales de cours ou échelons de cotation, ne sont pas, par essence, des variables continues mais discrètes. Cette caractéristique restreint les rendements des titres à prendre un nombre limité de valeurs. L'existence de cette structure particulière, qui est un *artefact* de la microstructure des marchés financiers, ouvre plusieurs voies de recherches, notamment sur les biais potentiels qu'elle cause dans les estimations des modèles. La caractérisation quantitative de ce phénomène a fait l'objet de quelques tentatives. Koppl et Nardone (2001) et Vorlow (2004) ont proposé des transformations des axes du diagramme de phases pour détecter le phénomène ; Wang et Wang (2002) et Mitchell et Mckenzie (2006) ont quant à eux utilisé l'alignement des droites. Nous nous inspirons de ces articles principaux pour proposer et définir trois indicateurs d'intensité du phénomène de rose des vents ($CRII_1$, $CRII_2$ et $CRII_3$). Ces mesures complémentaires et originales fondées sur plusieurs définitions du phénomène nous permettent de caractériser plus avant le phénomène de rose des vents.

Le premier indice d'intensité, appelé *Compass Rose Intensity Index 1* ($CRII_1$) repose sur l'idée suivante : le phénomène de rose des vents est caractérisé par la présence de plusieurs droites (passant toutes par l'origine) et reliant les différents couplets de rendements successifs. De ce fait, pour évaluer l'intensité du phénomène, il suffit de déterminer le nombre de droites de ce type ; ainsi, plus ce nombre est faible, plus les points sont bien alignés et plus le phénomène est visible. L'indice est ainsi défini à partir du rapport entre le nombre de droites constituant une rose des vents et le nombre de droites passant par l'origine. Théoriquement, si le phénomène de rose des vents existe dans la série considérée, l'indicateur doit tendre vers 100%. Dans le cas d'une rose des vents « parfaite », nous n'aurons que des droites passant par l'origine et rayonnant dans toutes les directions cardinales (droites bien placées). A l'opposé, plus le nombre de droites ne passant pas par les

points cardinaux (droites mal placées) sera important et moins le phénomène de rose des vents sera présent. Le phénomène est alors bien caractérisé par le nombre de droites « mal placées » ; d'où la formulation du premier indicateur qui est le rapport entre le nombre de droites dites « mal placées » et le nombre de points cardinaux et inter-cardinaux.

A la suite des travaux de Koppl et Nardone (2001) et plus tard de Vorlow (2004), nous avons construit le deuxième indice, appelé *Compass Rose Intensity Index 2* ($CRII_2$). L'idée sous-jacente est que la rose des vents peut être parfaitement définie par les angles du diagramme de phases des rendements successifs. A partir de ce constat, Koppl et Nardone (2001) proposent de transformer les couples de rendements en leurs coordonnées polaires. Ainsi, en comptabilisant les couples de rendements dont les secondes coordonnées polaires sont similaires, nous obtenons, d'après la terminologie de Koppl et Nardone (2001), un histogramme de *thêta* (*Empirical Theta Histogram*), mettant en lumière les distorsions liées à l'effet de discrétisation, puisqu'il permet d'amplifier le phénomène sous-jacent. Cet histogramme a pour but, de manière simple, de rendre compte de la dispersion des valeurs de la seconde coordonnée polaire normalisée et on peut comparer l'histogramme empirique à celui d'une loi uniforme pour mettre en évidence un effet de discrétisation. On peut aussi isoler le phénomène de rose des vents en comparant la distribution des coordonnées polaires de la série sous-jacente avec celle d'une pseudo-série obtenue en tirant au hasard des rendements dans la série originale (avec remise) et en les associant de manière aléatoire selon une loi uniforme (procédure de *bootstrapping*). Dans ce second cas de figure, nous pouvons isoler l'effet de corrélation de l'effet de discrétisation. S'appuyant sur ces considérations, l'indice d'intensité de rose des vents proposé est simplement défini par la somme des distances du khi-deux entre la fonction de densité empirique des secondes coordonnées polaires normalisées et la loi uniforme d'une part, et d'autre part, la fonction de densité empirique associée à la série de couples de rentabilités réarrangées aléatoirement. Ainsi l'indice prendra de grandes valeurs s'il existe un effet de discrétisation des prix, car la première partie de l'indice sera élevée ; l'indice sera aussi élevé lorsque la structure de dépendance temporelle de la série différera nettement de

l'hypothèse d'indépendance temporelle, ce qui contribuera à augmenter la seconde partie de l'indice.

L'idée du troisième indice, appelé *Compass Rose Intensity Index 3* ($CRII_3$) est la suivante : le phénomène de rose des vents se manifeste par la présence des couplets marqués (pixels) dans des parties spécifiques du diagramme de phases. Si l'on représente ce diagramme sous la forme d'un tableau en discrétisant les axes (tableau de contingence), alors certaines parties de ce tableau devraient être plus remplies que d'autres. Ainsi, en considérant tous les couples de rendements successifs, il est possible de construire un tableau de contingence. On peut obtenir une mesure similaire du phénomène directement à partir des rendements et non plus à partir de leurs coordonnées polaires. Là encore, on peut aussi essayer de distinguer l'effet de discrétisation de l'effet de dépendance temporelle en comparant le tableau de contingence empirique avec celui provenant d'une loi uniforme et issu de la série « bootstrappée ». Cet indicateur est cette fois-ci la distance mesurée par la norme de Frobenius calculée sur les différences des matrices de contingence. Cet indice aura de grandes valeurs s'il existe un effet de discrétisation des prix car la première partie de l'indice sera élevée ; l'indice sera aussi élevé quand la structure de dépendance temporelle de la série diffèrera nettement de l'hypothèse d'indépendance temporelle, ce qui contribuera à augmenter la seconde partie de l'indice.

Après avoir défini l'intensité du phénomène de rose des vents à partir des trois indices précédents, nous étudions les comportements de ceux-ci par des procédures déterministes et stochastiques à partir des données simulées et des données réelles de quelques titres vifs du marché français des actions. Nous mettons ainsi en évidence les conditions d'existence de ce phénomène, et par cette voie détournée, certains des éléments nécessaires à son apparition. Nous trouvons à partir des séries artificielles constituées de tous les rendements possibles évalués sur un intervalle de prix donné, que le phénomène n'apparaît pas de manière claire malgré la discrétisation. Par contre, sur des données simulées à partir de processus de rendements bien connus de la littérature financière avec plusieurs

spécifications et modifications de la structure auto-corrélative, nous observons de manière réelle ce phénomène. La structure auto-corrélative est donc une condition importante pour l'apparition du phénomène.

Nous trouvons donc que les principaux indicateurs proposés permettent - sauf cas pathologique - de bien caractériser le phénomène sur les données simulées. Nous avons aussi utilisé la base des données à très haute fréquence (données de transactions) de quelques titres vifs du marché français des actions. Nous avons ainsi défini un indice de classement des points qui confirme les conclusions avancées par les études précédentes. Les indices d'intensité proposés sont en effet nuls lorsque les prix des actifs sont continus ; de plus, les indices montrent qu'il existe une relation positive entre le phénomène de rose des vents et le ratio taille de l'échelon de cotation sur volatilité de la série des rendements. Nous avons aussi mis en évidence de manière simple le rôle de la taille de l'échantillon, et mis en avant l'hypothèse que ce sont les grandes variations de prix - peu présentes dans les petits échantillons - qui ont tendance à perturber la netteté du phénomène. Nous avons aussi montré que quelques processus utilisés dans la littérature financière sont tout à fait compatibles avec le phénomène de rose des vents ; autrement dit, qu'ils rendent compte suffisamment finement des caractéristiques financières - et en particulier de leur dépendance temporelle - pour ne pas être mis en défaut par ce phénomène de microstructure. Après avoir mis en évidence le phénomène sur le marché français en considérant quelques uns des titres les plus liquides de la place, nous avons enfin déterminé les seuils d'agrégation temporelle pour lesquels le phénomène tend à disparaître (à trente minutes, les effets de rose des vents biaisent moins les principales statistiques financières). Cette partie est à rapprocher des travaux de Gottlieb et Kalay (1985) et de Amilon et Byström (2000) qui s'interrogent sur les effets de la présence de ce type de phénomène sur les propriétés des estimateurs des paramètres caractéristiques de modèles de séries temporelles, et de Oomen (2002) et Griffin et Oomen (2008) qui montrent qu'il existe un seuil d'agrégation *optimal* pour mesurer la volatilité réalisée (*realized volatility*), en minimisant le biais de mesure lié à la dépendance des rendements de haute fréquence.

Les mesures proposées dans ce chapitre ont pour vocation principale d'aider à caractériser de manière plus formelle le phénomène de rose des vents *via* le calcul d'indicateurs quantitatifs simples. A cet égard, nous montrons par simulation et sur données réelles qu'ils caractérisent bien les situations de présence indéniable du phénomène. A cet effet, les résultats obtenus montrent que le phénomène a de très grandes chances d'être pleinement caractérisé si les trois indicateurs sont relativement élevés (supérieurs à 60), alors qu'au contraire, ceux-ci sont faibles (inférieurs à 20) lorsque l'existence du phénomène ne peut être invoquée. Entre ces deux limites extrêmes, le phénomène par définition existe, mais est plus ou moins visible. Les bornes mentionnées ci-dessus (existence du phénomène visible au-dessus du seuil de 20) sont similaires à celles proposées par Wang et Wang (2002) concernant leur mesure de visibilité de la structure. Alors que le phénomène est très présent en temps de transactions, l'on constate qu'après agrégation à une fréquence de trente minutes, le phénomène n'est plus caractérisé.

Partant des remarques et considérations générales de l'article séminal de Crack et Ledoit (1996), nous avons ainsi analysé le phénomène de rose des vents dans son ensemble. Nous avons d'abord montré que celui-ci apparaissait uniquement lorsque les prix des actifs étaient discrets et que la discrétisation des prix avait plus d'effet (relatif et absolu) sur les petites variations de prix que sur les grandes. Ensuite, nous avons constaté que le phénomène de rose des vents n'était visible qu'à partir d'un certain seuil du ratio taille de l'échelon de cotation sur volatilité de la série. Ce résultat confirme les conclusions de Crack et Ledoit (1996), Lee *et alii* (2005), McKenzie et Frino (2003), ou de Wang *et alii* (2000) qui montrent l'existence des seuils d'apparition du phénomène de rose des vents.

Notre objectif dans ce premier chapitre est de mieux caractériser et comprendre ce phénomène de rose des vents sur le marché français, en mettant en évidence certains facteurs explicatifs. Nous avons ainsi caractérisé le phénomène à l'aide de trois indicateurs complémentaires. Nous confirmons le fait qu'il résulte principalement d'un effet de

discrétisation et de corrélation, plus ou moins sensible en fonction de la volatilité, de la période d'échantillonnage et de la taille de l'échantillon et, enfin, que le phénomène est assez général sur le marché français. Ce chapitre a un intérêt pratique dans le cadre de l'utilisation de données de haute fréquence et de la justification du choix de la fréquence d'observation optimale. Ainsi l'intérêt de l'étude de ce phénomène, en dehors de son caractère de curiosité et de recherche de connaissances supplémentaires sur les données de haute fréquence, se trouve principalement dans le choix d'un seuil d'agrégation temporelle acceptable pour l'utilisation de modèles financiers sur des données de haute fréquence. Néanmoins, comme l'écrivent Mitchell et McKenzie (2006, page 3519), le phénomène ne revêt pas *a priori* « d'intérêt particulier pratique » direct en terme de gestion d'actifs. L'étude du phénomène de « rose des vents », au delà de son aspect de curiosité et de son caractère phénoménologique qui nous amène à nous intéresser aujourd'hui de plus en plus aux données de haute fréquence, permet toutefois aussi de s'intéresser indirectement à certains autres aspects pratiques liés au choix de la fréquence d'observation pertinente. Ainsi, le choix des seuils d'agrégation temporelle acceptables pour l'utilisation de modèles financiers sur des données de haute fréquence pourrait être guidé aussi par l'absence de phénomène. En effet, un courant récent de la littérature s'intéresse au contenu informationnel des données de haute fréquence (Cf. Engle et Russel, 1997).

Des grandeurs dites « réalisées », calculées à partir de données de haute fréquence, ont été proposées récemment comme la volatilité (Cf. Andersen *et alii*, 2000, 2001), les betas (Cf. Bollerslev et Zhang, 2003), ou encore la *Value-at-Risk* (Cf. Dacorogna *et alii*, 2001 ; Giot et Laurent, 2004 et Colletaz *et alii*, 2007) pour la mesure et le contrôle de risque des *traders*. Or, à très haute fréquence (données de transactions), il a été montré qu'il existait des biais microstructurels importants (Cf. Oomen, 2002 et Griffin et Oomen, 2008). Ainsi, il existe un compromis à trouver entre, d'une part, se rapprocher le plus possible du « vrai » processus des prix de marché en utilisant des données de haute fréquence et d'autre part, incorporer du bruit de microstructure dans le calcul d'un estimateur théoriquement plus précis (Cf. Aït-Sahalia *et alii*, 2005 et Nielsen et Frederiksen, 2008).

Il est donc intéressant de voir à partir de quel seuil d'agrégation temporelle le phénomène - très présent sur le marché français en temps de transactions - disparaît, et de définir ainsi une fréquence-type d'observation qui permette de réduire ce phénomène de bruit (Cf. Oomen, 2002 et Nielsen et Frederiksen, 2008). De manière plus « philosophique », nous pensons que ce phénomène de rose des vents nous invite aussi à nous interroger sur la qualité des estimations des modèles de finance empirique, en considérant que les propriétés des processus sous-jacents risquent d'être en partie perdues dans le phénomène de « rose des vents ».

Nous espérons étendre cette étude sur l'ensemble des places boursières composant l'*Euronext* (Amsterdam, Bruxelles, Lisbonne et Paris). Nous disposons désormais d'une base de données de transactions provenant d'*Euronext*, allant du 1^{er} janvier 2002 au 31 décembre 2007 contenant tous les titres. Nous pourrions ainsi tester la présence du phénomène sur les autres marchés en tenant compte des mouvements récents des prix des actifs. Les principales valeurs limites des indicateurs de roses des vents, déterminées sur quelques titres du marché français au premier semestre 2004, pourront ainsi être mises à jour et généralisées à l'échelle européenne. Nous envisageons aussi de faire une analyse comparative des différentes mesures de visibilité du phénomène de rose des vents disponibles dans la littérature (Cf. Koppl et Nardone, 2001 ; Wang et Wang, 2002 et Mitchell et McKenzie, 2006). Comme prolongement naturel de ce chapitre, il serait intéressant de voir comment les paramètres des modèles *GARCH* se comportent de manière empirique par simulations sur des données continues (sans pas d'échelon de cotation) et sur des données discrétisées avec plusieurs pas d'échelon de cotation différents. Nous pourrions alors à partir des indicateurs de rose des vents, déterminer l'ampleur des erreurs d'estimation engendrées par la discrétisation et la structure de corrélation imposées. Un point de départ de cette étude concerne les travaux de Amilon et Byström (2000) et de Amilon (2003), qui montrent comment le phénomène biaise les principaux tests statistiques (voir en annexe les articles en cours de finition).

Résumé du chapitre sur “*High-frequency Market Risk Scenarii by Independent Component Analysis: Method and Application*”

A la suite des multiples crises financières de ces dernières décennies, les institutions financières ont proposé des cadres réglementaires et des indicateurs pour prévenir les crises et leurs effets. Pour cela, l'accent a été particulièrement mis sur la mesure des risques non-diversifiables et ceci dans la nécessité de bien comprendre les possibles implications sur la valeur des actifs (sévérité du risque) et la fréquence d'occurrence de ce risque (probabilité). Cette réalité a créé le besoin d'une mesure regroupant à la fois la probabilité et l'impact potentiel d'un certain risque. La Valeur-en-Risque (*VaR* en anglais pour *Value-at-Risk*), qui intègre ces deux dimensions de l'analyse, est un indicateur très répandu sur les marchés financiers, pour des nécessités réglementaires et aussi pour l'amélioration de la gestion des portefeuilles en offrant une mesure de risque complète et fiable.

La *VaR* se définit comme la perte maximale attendue à un horizon temporel donné associé à un certain niveau de confiance. Elle est employée pour la gestion des risques, y compris les risques de crédit et les risques opérationnels (Duffee et Pan, 1997 et 2001 et Tsai, 2004). L'objectif du calcul de la *VaR* est de fournir une évaluation dynamique et fiable du risque d'un actif ou d'un portefeuille. Pour atteindre cet objectif, le cadre doit être flexible et assez général pour s'adapter à des changements de conjoncture dus aux variations des facteurs de risque associés. Le concept de *VaR* a été popularisé par les amendements du comité de Bâle de 1996 sur les risques de marché dans le cadre du capital de normalisation. Les agences de surveillance autorisent aux banques d'employer leurs propres modèles internes de *VaR* pour calculer leurs besoins en capitaux de normalisation comme alternative à la méthodologie conventionnelle. Depuis, cet indicateur est devenu très utile pour le contrôle interne de chaque banque et des gérants de portefeuilles. Celui-ci offre en effet une possibilité supplémentaire d'évaluer et de maîtriser l'exposition des

placements aux aléas de la conjoncture économique ou des marchés. En dépit de sévères critiques sur la fiabilité de la mesure, l'utilisation systématique de la *VaR* exerce un effet positif global sur la gestion du risque dans les institutions financières, intégrant ainsi le degré croissant de complexité des instruments d'investissement au cours des dernières années. Sur le plan académique, depuis les propositions du cadre réglementaire du Comité de Bâle, la littérature financière sur les mesures et l'évaluation des risques financiers s'est accentuée.

Deux grandes familles de modèles ont été proposées : la *VaR* paramétrique et la *VaR* non-paramétrique. La *VaR* paramétrique est une mesure de risque basée sur une hypothèse distributionnelle. En effet, cette méthode suppose que le gestionnaire de portefeuille est en mesure d'estimer les paramètres de la distribution des rendements générés par un titre ou par un portefeuille de titres. Couramment, la distribution normale est utilisée pour modéliser les rendements d'un actif. Ainsi, la *VaR* est calculée à partir des deux premiers moments de la distribution des rendements. Dans ce cas, la *VaR* est en effet parfaitement définie par la moyenne et l'écart-type de la distribution. Par contre, les méthodes non-paramétriques et historiques ne nécessitant aucune hypothèse sur la distribution des rendements, les calculs sont directement effectués sur les données observées. Celles-ci permettent de déterminer par classement la distribution des pertes possibles et leurs fréquences d'apparition en fixant le niveau de percentile voulu (1%, 2%, 5%, ...).

Cette mesure de risque, par définition, est bien adaptée pour les actifs unitaires avec des profils de risque homogènes. Mais, lorsqu'un portefeuille d'actifs est considéré, il est difficile d'obtenir une agrégation du risque de ce portefeuille sans perdre les spécificités de chaque ligne. En effet, le risque global du portefeuille n'est pas simplement une agrégation des risques de chaque actif composant le portefeuille. Pour résoudre ce problème et calculer de manière efficace les risques associés aux actifs sur un marché, les modèles traditionnels en finance supposent l'existence de facteurs globaux qui influencent l'évolution des prix sur le marché (*Cf.* Ross, 1976 et Gençay *et alii*, 2003). En effet, plusieurs fac-

teurs interagissent pour modifier les mouvements des rendements des actifs, entraînant des évolutions plutôt complexes et imprévisibles. Il n'est donc pas aberrant de supposer que les mouvements des différents prix soient dirigés par des facteurs communs avec des caractéristiques spécifiques. Loretan (1997) exploite cette idée en extrayant les principaux facteurs statistiques par une Analyse en Composantes Principales (ACP) sur les rendements des actifs. L'Analyse en Composantes Principales est une technique statistique qui permet d'agréger les facteurs en groupes orthogonaux tout en maintenant autant que possible la variabilité des données originales. Cette technique permet d'atteindre deux objectifs de simplification de données : la réduction du nombre de facteurs et la détection des relations entre ces variables. Les facteurs communs ainsi identifiés sur des données du marché permettent de générer et d'agréger facilement des trajectoires virtuelles dans le temps (à cause de l'homogénéité des facteurs et des corrélations nulles entre ceux-ci).

Ces méthodes de calcul de VaR sont le plus souvent utilisées pour des fréquences d'observation journalières, hebdomadaires, mensuelles voire annuelles. Mais, étant donné la disponibilité de nouvelles bases de données plus complètes (données de transactions), une littérature récente propose des modèles de gestion des risques financiers utilisant des données intra-journalières pour des mesures de risque plus sensibles aux mouvements des marchés (Cf. Dacorogna *et alii*, 2001 et Rydberg et Shephard, 2003), et pour une meilleure évaluation et allocation de la valeur du portefeuille (Cf. Bollerslev et Zhang, 2003). Ces données de haute fréquence présentent l'avantage d'être proches du « vrai » processus générateur des prix. Toutefois, comme nous l'avons vu au chapitre précédent, il est nécessaire d'utiliser des méthodes ou techniques qui tiennent compte des biais microstructurels observés. De plus, la non-linéarité et la non-gaussianité sont des caractéristiques spécifiques de ce type de données (Cf. McNeil et Frey, 2000 ; Härdle *et alii*, 2003 et Voit, 2003). De ce fait, les deux premiers moments ne peuvent plus suffire à caractériser les facteurs qui influencent les mouvements de prix. La technique d'Analyse en Composantes Principales, qui nécessite un filtre linéaire pour la détection des facteurs communs, entraîne ainsi une perte d'information lorsque des données financières à haute

fréquence sont considérées. Dans ce cas, une Analyse en Composantes Indépendantes (*Independent Component Analysis - ICA* - en anglais, Cf. Comon, 1994 et Hyvärinen *et alii*, 2001) semble être une technique plus appropriée comme cela est mentionné par Back et Weigend (1997).

L'Analyse en Composantes Indépendantes est une technique d'extraction de données multidimensionnelles qui permet de trouver la structure latente dans les données à partir des moments d'ordre supérieur. Cette méthode permet ainsi de dépasser le seul cadre moyenne-variance de l'Analyse en Composantes Principales en considérant pour les estimations des sources inconnues, les moments d'ordre supérieur. La méthode d'Analyse en Composantes Indépendantes (ACI) qui sera brièvement présentée dans la suite de ce rapport, est bien connue dans le traitement des signaux et est appliquée dans divers domaines (signaux de biomédecine, signaux de télécommunications, . . .). Un signal peut être vu comme un mélange de différentes sources physiques et de sources extérieures de bruits. L'exercice est de trouver les sources indépendantes qui forment au final le signal. Partant de cette technique plus complexe et originale (extraction de facteurs et spécification de *scenarii* plus réalistes tout en conservant la structure des données), nous nous proposons de mesurer de façon plus robuste la *VaR* d'un portefeuille de titres financiers. Avant de présenter notre procédure de calcul de la *VaR*, nous revenons sur la technique de l'Analyse en Composantes Indépendantes.

L'Analyse en Composantes Indépendantes (Cf. Comon, 1994 et Hyvärinen *et alii*, 2001) est une méthode statistique qui exprime un ensemble d'observations multidimensionnelles comme combinaison de variables latentes inconnues. On considère que ces variables latentes sont des sources (ou composantes indépendantes) et sont statistiquement indépendantes les unes des autres par construction. Cette définition fait ressortir les caractéristiques importantes de cette technique (prises en compte lors des estimations). En effet, tout d'abord, les sources sont supposées statistiquement indépendantes. Cette hypothèse est justifiée par le fait que les facteurs ou sources de risques qui influencent

les cours des actifs peuvent être d’origines très différentes (politique étrangère, variables micro-économiques ou macro-économiques, ...). D’autre part, le modèle suppose qu’une seule source au plus a une distribution gaussienne. Cette hypothèse trouve ici sa place dans la mesure où, avec les données financières de haute fréquence, les signaux normalement distribués et indépendants sont très rares.

Plusieurs approches pour estimer les composantes indépendantes ont été proposées dans la littérature, fondées sur plusieurs familles d’algorithmes (*Cf.* Cardoso et Souloumiac, 1993 ; Cardoso, 1999 et Hyvärinen *et alii*, 2001). Dans toutes les approches, une fonction objective (ou fonction contraste) est d’abord choisie. Cette fonction qui est non-linéaire, permet d’évaluer à chaque étape d’optimisation la qualité des résultats.

Les premières approches d’estimation des composantes indépendantes sont basées sur des méthodes tensorielles. Les algorithmes les plus connus sont les *FOBI* (pour *First-Order Blind Identification* ; *Cf.* Cardoso, 1989 et 1990) et *JADE* (pour *Joint Approximate Diagonalization of Eigenmatrices*, *Cf.* Cardoso et Souloumiac, 1993). Ces méthodes utilisent des opérateurs mathématiques comme les tenseurs et les moments. Les tenseurs sont des généralisations des opérateurs linéaires. Les moments d’ordre supérieur sont aussi utilisés pour les estimations des composantes (Hyvärinen *et alii*, 2001).

La deuxième approche pour l’évaluation du modèle d’ACI est l’estimation par le maximum de vraisemblance (*Cf.* Bell et Sejnowski, 1995). Dans cette seconde famille, on choisit les valeurs de paramètres qui donnent les plus hautes probabilités aux observations communes.

La troisième approche d’estimation est basée sur la maximisation directe de la non-gaussianité des composantes. La justification théorique de cette maximisation vient de la propriété du théorème centrale limite qui stipule que toute combinaison linéaire de variables indépendantes non gaussiennes est « plus gaussienne » que les variables aléatoires

d'origines. Pour mesurer la gaussianité, plusieurs statistiques sont proposées (la *kurtosis*, l'entropie, ou les *cumulants* d'ordre supérieur). Mais, ces mesures qui nécessitent de connaître la distribution de probabilité sont très sensibles aux bruits et aux valeurs aberrantes. En conséquence, d'autres approximations et améliorations ont été proposées par Hyvärinen et Oja (1997) ; Hyvärinen (1999) et Hyvärinen *et alii* (2001). Ces améliorations concernent principalement les mesures du caractère non-gaussien. Un algorithme qui peut être employé dans toutes les approches précédemment énumérées d'ACI est le *FastICA* développé par Hyvärinen *et alii* (2001). C'est un algorithme à point fixe itératif avec une approximation de la *kurtosis* par une fonction non linéaire basée sur les densités empiriques. C'est cette dernière approche, rapide et robuste en pratique, que nous avons choisi pour nos estimations des facteurs de risque pour le calcul de la *VaR*.

L'utilisation de cette technique d'ACI vient du fait qu'un signal global (observable) provoqué par un ensemble de signaux est une superposition des composants élémentaires (non observables) qui se produisent indépendamment. La technique est différente de l'Analyse en Composantes Principales (ACP) dans le sens où elle impose non seulement la décorrélation des composantes, mais également l'indépendance calculée sur les moments d'ordre supérieur. Par ailleurs, Back et Weigend (1997) ont montré que l'Analyse en Composantes Indépendantes domine l'Analyse en Composantes Principales dans un contexte financier. En analysant la série chronologique avec l'ACI, les facteurs de risques ou composantes fournis sont plus en adéquation avec la situation réelle sur les marchés financiers. La décomposition de facteurs statistiques est validée avec une interprétation économique. Les facteurs cachés indiqués dans l'ACI pourraient inclure des nouvelles informations (intervention du gouvernement, catastrophes naturelles, événements politiques, ...), des réponses aux grands mouvements des marchés ou des bruits complexes non expliqués (Lai *et alii*, 1999 ; Moody et Yang, 1999 ; Cha et Chan, 2000 ; Glasserman *et alii*, 2002 ; Cheung et Xu, 2001 et Rojas et Moody, 2001). Comment exploiter alors cette technique originale d'extraction de sources de risque pour l'évaluation et la gestion des risques financiers ?

La *VaR* est un concept relativement simple, ce qui permet son utilisation par l'ensemble des gestionnaires. En revanche, afin d'obtenir un bon estimateur (d'un point de vue méthodologique), la *VaR* peut faire appel à des notions avancées de statistiques stochastiques et de probabilités. La modélisation de la distribution de probabilité devrait faire l'objet d'une attention particulière. Le rendement du portefeuille dépendant des variations de plusieurs sources indépendantes, nous utilisons l'extraction par Analyse en Composantes Indépendantes des facteurs de risques. Concrètement, en modélisant les rendements d'un portefeuille en fonction des rendements de chaque ligne, le portefeuille sera ainsi relié directement aux sources qui influencent les titres. Il suffit alors de déterminer et de bien modéliser chaque source de risque pour mieux évaluer le portefeuille. Dans les modèles traditionnels, les seules restrictions concernent les moments d'ordre 2 qui sont nuls (extraction par l'Analyse en Composantes Principales). Mais, nous savons que la décorrélation n'entraîne pas forcément l'indépendance. Par conséquent, la seule restriction typique sur des facteurs décorrélés dans les modèles à facteurs ne peut garantir que la valeur du portefeuille soit liée à l'influence d'autres facteurs extrêmes avec des distributions à queues épaisses. Au contraire, si les facteurs sont indépendants entre eux, il est possible de construire un tel portefeuille.

Loretan (1997) emploie une ACP pour extraire des facteurs non-corrélés dans le cadre du calcul de la *VaR*; Albanese *et alii* (2002) utilisent plutôt des approximations pour réduire la dimension du facteur, tandis que nous proposons ici d'employer la méthode d'ACI pour l'extraction des facteurs indépendants (*Cf.* Back et Trappenberg, 1999 et Lai *et alii*, 1999). L'ACI intervient donc dans l'estimation des composantes et pour chaque facteur déterminé, nous modélisons ses moments par des distributions distinctes pour mieux apprécier les divergences dans les mesures proposées. Nous avons adapté les modèles paramétriques aux signaux indépendants de l'ACI ou aux facteurs non-corrélés de l'ACP. Puisque nous devons capturer les événements extrêmes sur les queues de la distribution des rendements du portefeuille (*Cf.* Gonzalo et Olmo, 2004), une distribution à queues

épaisses, telle que la distribution de *Pearson Type IV* ou la distribution de *t-Location Scale* sont employées. Ces distributions ont l'avantage d'intégrer les moments d'ordres 3 et 4 dans leur spécification. La distribution gaussienne s'ajoute également à la modélisation des facteurs prédéterminés par l'ACP et par l'ACI pour avoir une bonne lisibilité de la performance des mesures de risques associées.

Plusieurs mesures d'évaluation du risque d'un portefeuille à partir des distributions homogènes proposées dans la littérature (*Cf.* Back et Weigend, 1997 et Giot et Laurent, 2004) sont utilisées pour comparer les différentes approches. Il s'agit de la *VaR* gaussienne (caractérisée par la moyenne et l'écart-type des rendements du portefeuille), de la *VaR* Historique (percentile des bénéfices/pertes historiques du portefeuille), de la *VaR* Risk-Metrics (corrige la volatilité des rendements par une moyenne exponentielle) et la *Student-VaR* (les rendements du portefeuille suivent des distributions de *Student*). Les mesures de *VaR* appliquées sur des données issues d'une Analyse en Composantes Indépendantes est alors appelée *ICA-VaR*.

Pour comparer et analyser la fiabilité des différentes mesures alternatives de *VaR* proposées, plusieurs procédures de *backtesting* ont été implémentées sur la base des tests les plus utilisés dans la littérature. Nous avons utilisé le test de proportion d'échec (Kupiec, 1995) qui est basé sur l'estimation du nombre de dépassements de la *VaR* estimé par le modèle. Sous l'hypothèse nulle, les prévisions du modèle *VaR* sont en moyenne dans l'intervalle de significativité. Nous avons aussi utilisé le test de l'intervalle de confiance de Christoffersen (1998) qui est lié au ratio de vraisemblance ; il teste conjointement si la proportion d'échecs est conforme à ce qui est anticipé et si les exceptions (échecs dans la prévision) sont indépendantes les unes des autres. L'autre méthode de *backtesting* utilisée dans ce chapitre est le test du quantile dynamique de Engle et Manganelli (2004). L'idée de base de ce test est qu'un bon modèle doit produire un ensemble indépendant d'échecs. Sous l'hypothèse nulle, tous les coefficients de la régression du modèle avec constante sont nuls.

En utilisant des données de prix *intraday* de janvier 2002 jusqu'à décembre 2006 pour 43 titres vifs les plus liquides du marché français des actions, nous appliquons ces procédures de calcul du risque sur le portefeuille équi-pondéré. Les valeurs des *VaR* obtenues pour l'ensemble des méthodes utilisées sont plutôt homogènes et relativement faibles (moins de 1% de perte possible pour tous les niveaux de confiance). De même, les rendements du portefeuille qui dépassent le niveau de la *VaR* (*CVaR*, pour *Conditional Value-at-Risk* en anglais) ont une moyenne inférieure à 0,01%. Ces résultats sont tout à fait conformes aux données de haute fréquence qui ont des rendements relativement bas. Toutefois, le niveau de la *VaR*, qui est calculé sur une période fixe, ne fournit pas d'informations réelles sur la qualité des mesures. Il faut donc recourir aux calculs dynamiques pour apprécier le comportement de chacune des méthodes.

Les différentes mesures de *VaR* sont validées par des tests de spécification les plus utilisés dans la littérature pour le contrôle des mesures de risque (Christoffersen, 1998 ; Christoffersen et Pelletier, 2003 et Ferreira et Lopez, 2005). En effet, les *backtests* appliqués sur les différents modèles montrent qu'en général les résultats sont acceptables pour la plupart des méthodes de calcul de *VaR* présentées dans ce chapitre. Toutefois, nous observons que la *VaR* basée sur l'ACI avec la distribution de *Pearson Type IV* fournit la meilleure estimation du risque du portefeuille pour plusieurs niveaux de confiance. La distribution du nombre d'échecs est en effet proche de la valeur théorique des degrés de confiance considérés pour l'*ICA-VaR*. Ce résultat confirme que les méthodes classiques d'évaluation de la *VaR* sont améliorées par une meilleure utilisation des données de haute fréquence. L'*ICA-VaR* proposée réagit mieux aux mouvements de marché observés à très haute fréquence et rend bien compte de la réalité des événements. Ainsi, l'approche basée sur l'Analyse en Composantes Indépendantes présente donc des atouts pour une meilleure prise en compte de ces caractéristiques intrinsèques des données de haute fréquence. Ces résultats sont à rapprocher de ceux de Giot et Laurent (2004) qui montrent, par une modélisation des dépendances par des modèles à *duration*, qu'à haute

fréquence les modèles classiques ne permettent pas de prédire de manière fiable la *VaR*.

Nous montrons ainsi que dans un environnement caractérisé par la non-gaussianité et de fortes non-linéarités, la méthodologie de l'ACI peut, par une exploitation appropriée, révéler la structure sous-jacente dans les données *intraday* qui sont nécessaires à la détermination des mesures de risque. Nous trouvons aussi, par plusieurs spécifications de *VaR* différentes et des *backtests*, que la *VaR* basée sur l'Analyse en Composantes Indépendantes, avec prise en compte des queues de distribution, caractérise mieux le risque du portefeuille. De façon intéressante, les « bruits » que nous observons dans les données *intraday* peuvent affecter les mesures traditionnelles. De ce fait, le calcul de la *VaR* avec les données dé-bruitées est plus robuste que la *VaR* déduite directement des données brutes ou de l'extraction des facteurs par l'ACP. Finalement, basée sur nos premiers résultats, la technique de l'ACI pourrait avoir un certain intérêt pour d'autres applications financières, comme le *stress test* ou bien l'allocation d'actifs.

L'utilisation de l'Analyse en Composantes Indépendantes pour la détermination des facteurs de risque offre de nouvelles pistes pour les modèles factoriels en général et pour l'évaluation des mesures de risque en particulier. Les résultats obtenus sur le calcul de la *VaR* en haute fréquence montre bien que cette technique est une alternative à l'Analyse en Composantes Principales dans les cas où l'hypothèse de distribution gaussienne des rendements n'est que trop approximative. Quelques améliorations et extensions sont toutefois en cours de traitement. La première extension provient de la modélisation des facteurs de risque. Dans des travaux en cours, nous testons sur les facteurs issus de l'ACI, des familles de distributions intégrant davantage les moments d'ordre supérieur ; notamment à partir des modèles *Skew-student* (Keith et Shen, 2004 et Adcock et Meade, 2007). Une autre piste de travail est de tester d'autres approches d'extractions des facteurs indépendants comme l'utilisation des copules (*COPICA* pour *Copula Independent Component Analysis* en anglais, Cf. Chen *et alii*, 2008) ou bien des méthodes avec contraintes

(Lu et Rajapakse, 2000 et 2005). En effet, la version de l'ACI présentée dans ce chapitre ne permet pas, comme dans le cas d'une ACP, de réduire le nombre de facteurs ; elle détermine automatiquement le nombre de facteurs statistiquement indépendants. Nous pouvons à partir d'une Analyse en Composantes Indépendantes avec contraintes (*cICA* pour *constraint Independent Component Analysis* en anglais, Cf. Lu et Rajapakse, 2000 et 2005), réduire la dimension des données tout en gardant le caractère d'indépendance des facteurs. De plus, nous pouvons supposer que les données observées sont des combinaisons des sources indépendantes qui sont affectées par des bruits. A cet effet, Chang et Zhang (2006) et Kopriva et Seršić (2008) ont proposé des algorithmes qui permettent de filtrer les bruits par la méthode des ondelettes (*Wavelet* en anglais) avant de déterminer les composantes indépendantes. Cette forme d'extraction de facteurs indépendants (*SDICA* pour *Subband Decomposition Independent Component Analysis* et *WP-SDICA* pour *Wavelet Packets-SDICA* en anglais) fournit directement la bande dé-bruîtée de fréquences dans laquelle les sous-composantes du signal d'origine sont indépendantes. Cette dernière technique applique un filtre aux observations pour permettre de décomposer l'échantillon en bandes de sous-composantes dépendantes et indépendantes. De même, une estimation de la performance des algorithmes à partir des critères spécifiques de l'ACI (*Amari's error* et *Comon Criteria*, Cf. respectivement Amari *et alii*, 1996 et Comon, 1994) depuis janvier 2002 est envisageable. Enfin, une étude comparative, plus générale, avec une synthèse des principales mesures de risque est aussi envisagée sur les titres européens les plus liquides à partir des données de transactions de l'ensemble des places boursières de l'*Euronext*.

Résumé du chapitre sur “*A Robust Conditional Multi-moment Asset Pricing Model with High-frequency Data*”

Dans ce troisième chapitre, nous abordons le cœur de la théorie financière moderne de la gestion d’actifs, en généralisant les approches traditionnelles et en complétant leurs analyses grâce à la connaissance de l’information contenue dans les données de haute fréquence.

Le modèle d’équilibre des marchés financiers (*Capital Asset Pricing Model, CAPM*) de William Sharpe (1964) et John Lintner (1965) marque la naissance de la théorie de l’évaluation d’actifs. Il offre de manière théorique de puissants outils de prévision pour la valeur d’un actif et fournit une relation supposée stable entre les rendements espérés et le risque prévu. Ce modèle d’évaluation des actifs financiers stipule que les rendements espérés des actifs sont des fonctions linéaires des covariances et des facteurs de risque liés au marché (risques systématiques). Sharpe (1964), Lintner (1965), Merton (1973), Ross (1976), Breeden (1979), Barone (1985) ou Dittmar (2002) ont proposé plusieurs formulations de ce paradigme général. Cependant, la plupart des tests empiriques proposés jusqu’ici ont produit des résultats négatifs ou ambigus. En effet, Shanken (1992) et Kan et Zhang (1999a et 1999b) ont montré, en utilisant des méthodologies de tests plus robustes, que les résultats pouvaient varier sensiblement en fonction du modèle utilisé. Ceci est dû, d’une part, aux choix de modèles incomplets et des processus d’estimation non conformes pour les tests empiriques, et d’autre part, aux faibles seuils de tolérance souvent utilisés, entraînant ainsi un manque de puissance (*Cf.* Kan et Zhou, 1999 et Jagannathan et Wang, 2002). Ces derniers notent en plus que ces choix peuvent mener, non seulement au rejet des modèles corrects, mais également à l’acceptation des sources de risques systématiques non pertinentes.

Pour améliorer la qualité des modélisations et mettre fin aux irrégularités empiriques sur les modèles classiques, un certain nombre de facteurs supplémentaires ont été détectés dans la littérature. Banz (1981) tout d'abord, trouve qu'il existe une relation significative entre les rendements des actions des sociétés et la taille ; plus tard, Fama et French (1993) mettent en évidence un lien entre le rendement attendu, la taille de la firme et le ratio de la valeur réelle de la firme sur la valeur boursière. Bien que la véracité de ces facteurs supplémentaires soit toujours discutée, il est évident que le simple cadre moyenne-variance du modèle CAPM est insuffisant pour caractériser les comportements réels des agents sur les marchés. Toutes les contributions précédentes considèrent de plus, seulement une stratégie moyenne-variance, l'investisseur construit son portefeuille avec des actifs risqués et l'actif sans risque. Or le critère moyenne-variance suppose implicitement que les rendements sont normaux et que les moments les plus élevés (au delà de la moyenne et de la variance) ne sont pas appropriés pour l'allocation d'actifs. Ceci, parce que les variables financières montrent le plus souvent de fortes particularités dans les distributions. Notamment, les rendements ont des queues de distribution épaisses, sont asymétriques et sont marqués par des phénomènes de regroupement (*clustering*). Il est donc important de prendre en compte ces caractéristiques par des modèles plus complets et des méthodologies d'estimation robustes et adaptées.

Kraus et Litzenberger (1976), Barone (1985) et Harvey et Siddique (2000), Barone *et alii* (2003) entre autres, ont, depuis lors, proposé des modèles d'évaluation d'actifs alternatifs non normaux en intégrant un facteur non linéaire mesurant l'asymétrie de la distribution des rendements (la *coskewness*). Kraus et Litzenberger (1976) et Harvey et Siddique (2000) stipulent que les rendements espérés sont fonctions non seulement de la covariance et des facteurs liés à la firme, mais aussi de la *coskewness* du portefeuille de marché. En particulier, Harvey et Siddique (2000) font ressortir l'importance de la *coskewness* sur l'augmentation du pouvoir explicatif des régressions en coupe transversale. Plus récemment, Dittmar (2002) et Jurczenko et Maillet (2006) ont présenté un cadre dans lequel les agents sont, au delà des facteurs précédents, sensibles à la *kurtosis*, impli-

quant que les rendements des actifs dépendent aussi de la *coskewness* et de la *cokurtosis* du portefeuille de marché.

Le pouvoir explicatif empirique des modèles avec extensions aux moments d'ordre supérieur semble plus élevé que celui des modèles classiques mais reste néanmoins faible sur des données de basse fréquence (journalières, hebdomadaires, mensuelles, ...). Parallèlement à cela, de par la disponibilité récente de bases de données sur certains marchés financiers, plusieurs études se sont intéressées aux caractéristiques statistiques des données de haute fréquence (voir par exemple Kunitomo, 1992 ; Bai *et alii*, 2000 ; Dacorogna *et alii*, 2001 ; Corsi *et alii*, 2001 ; Oomen, 2002 et Voit, 2003), et aux estimations des paramètres des modèles de basse fréquence à partir des données de haute fréquence (Andersen *et alii*, 2002 ; Barndorff-Nielsen et Shephard, 2002a ; Bollerslev et Zhang, 2003 et Brandt et Diebold, 2004), ce choix des données de haute fréquence étant motivé par le fait qu'elles sont proches des « vrais » processus générateurs des prix (*Cf.* Kunitomo, 1992).

La première statistique empirique étudiée sur les modèles de basse fréquence, dans la littérature économétrique financière, a été le moment d'ordre deux (la volatilité, *Cf.* Andersen *et alii*, 2002 ; Barndorff-Nielsen et Shephard, 2002a et Zumbach *et alii*, 2002). La volatilité instantanée d'une journée est alors approximée par la somme des rentabilités au carré à l'intérieur d'une journée (volatilités réalisées). Cette approche, basée sur les données de haute fréquence, fournit à l'évidence des possibilités très larges pour l'étude des fluctuations et leur persistance dans les modèles conditionnels, et aussi pour les covariances conditionnelles des actifs individuels. Bollerslev et Zhang (2003) s'inspirent de cette définition de la volatilité dite « réalisée » pour proposer une estimation des bêtas « réalisés » avec des données *intraday*. Ils montrent que les mesures et les co-mesures « réalisées » sont plus précises pour l'évaluation des risques systématiques dans les modèles à facteurs par rapport aux modèles classiques. Mais jusqu'à présent, aucune étude n'a encore proposé une extension des mesures réalisées (volatilités réalisées et bêtas réalisés) aux moments réalisés d'ordre supérieur.

Ainsi, partant de l'approche standard de Bollerslev et Zhang (2003), nous incluons dans ces modèles d'évaluation d'actifs des moments réalisés d'ordre supérieur. En effet, nous nous inspirons des justifications théoriques de Jurczenko et Maillet (2006) pour introduire dans les modèles conditionnels d'évaluation (*Cf.* Bollerslev *et alii*, 1988 et Jondeau et Rockinger, 2004), les moments d'ordres trois et quatre. La motivation pour l'évaluation conditionnelle d'actifs en multi-moment vient du fait que, théoriquement, les considérations économiques financières suggèrent que les bêtas puissent varier avec des variables exogènes caractérisant les comportements des agents. Pour mieux exploiter ces informations apportées par les extensions proposées sur les modèles classiques - modèles d'évaluation avec des moments d'ordre supérieur (*Cf.* Jurczenko et Maillet, 2006), les estimations ont été menées avec des méthodologies qui permettent de neutraliser les erreurs de spécification et de mesure des données (*Cf.* Scholes et Williams, 1977 ; Berkowitz et Diebold, 1998 et Ledoit et Wolf, 2003 et 2004), tout en tenant compte des inter-relations entre les actifs financiers en terme de rentabilité (méthodes d'équations simultanées proposées par Zellner, 1962 ; et correction de Parks, 1967).

En effet, les estimations robustes du modèle proposé combinent les régressions apparemment indépendantes (*SUR* pour *Seemingly Unrelated Regression* en anglais) de Zellner (1962) et la méthodologie en coupe transversale de Parks (1967). Les régressions apparemment indépendantes permettent de tenir compte, non seulement des inter-relations entre les actifs et le portefeuille de marché, mais aussi des fortes corrélations entre les actifs entre-eux (par les corrélations inter-équations). Cette méthode d'estimation est ainsi plus robuste qu'une simple régression équation par équation. Nous complétons les régressions de Zellner (1962) par la méthode de correction de Parks (1967). Ceci parce que les régressions apparemment indépendantes (*SUR*), proposées par Zellner (1962), peuvent être analysées comme des cas particuliers des modèles généralisés avec certaines restrictions sur les corrélations inter-actifs (par exemple, modèles hétéroscédastiques ou de corrélations périodiques). Parks (1967) propose un procédé d'évaluation qui permet

d'intégrer cette dimension des interdépendances en coupe transversale, en corrigeant les valeurs estimées des principaux éléments de la régression en coupe longitudinale par une matrice de variance complémentaire.

Au delà des estimations robustes des modèles, nous avons effectué des contrôles et des tests de spécifications par des procédures bien connues de la littérature financière. Le premier groupe de test du modèle d'équilibre des marchés financiers est lié à la variation chronologique des séries (*Cf.* Black *et alii*, 1972 et Gibbons *et alii*, 1989) et le deuxième groupe est lié aux régressions en coupe transversale (Fama et Macbeth, 1973).

Black *et alii* (1972) ont suggéré une stratégie pour l'estimation et les tests des modèles d'évaluations d'actifs financiers. Ils proposent de régresser, par une méthode standard (Moindres Carrés Ordinaires, MCO), la prime de risque de chaque actif sur la prime de risque du portefeuille de marché, et de vérifier que la constante de régression de tous les actifs est nulle. Ceci peut être fait par un test du Khi-deux sur les paramètres estimés individuellement ou bien par un test du Khi-deux joint (*Cf.* Campbell *et alii*, 1997 et Cochrane, 2001). Toutefois, pour le test de Black *et alii* (1972) présenté précédemment, comme les variances réelles sont inconnues, une approximation de la variance par la théorie de la répartition de grands échantillons (Campbell *et alii*, 1997 et Cochrane, 2001) est nécessaire. Gibbons *et alii* (1989) améliorent ce test en donnant une distribution finie de l'échantillon pour les tests.

Contrairement aux deux approches chronologiques directes précédentes, Fama et Macbeth (1973) proposent un test en coupe transversale. Le CAPM impose une restriction, à savoir que les rendements sont linéairement liés au bêta du marché. De ce fait, le CAPM peut être spécifié par une régression en coupe transversale des rendements prévus sur des bêtas de chaque actif. Ainsi, pour examiner les modèles alternatifs, il suffit d'ajouter des variables explicatives supplémentaires à la régression et de tester la significativité des paramètres et la validité du modèle. Ce test a l'avantage de mettre en évidence la

relation qui lie le rendement espéré au bêta de l'actif sur un marché et il permet en outre d'évaluer l'influence de variables externes.

A partir de données de haute fréquence de 43 titres du marché français des actions et enregistrées toutes les 5 minutes du 2 janvier 2002 au 29 décembre 2006, nous avons construit les séries de style en suivant l'approche de Fama et French (1993) et de Bollerslev et Zhang (2003) afin de tester et comparer plus efficacement les différents modèles alternatifs. Les rendements relatifs des portefeuilles sont définis par le différentiel de rendements entre les portefeuilles regroupés selon les différentes valeurs *book-to-market* individuelles du titre et leur capitalisation boursière.

Bien que les titres les plus liquides dans notre échantillon soient généralement plusieurs fois échangés dans chaque intervalle de 5 minutes, certains titres, moins liquides, peuvent ne pas l'être pendant une heure ou plus. Ainsi, les rendements des portefeuilles sont clairement susceptibles de subir des effets non synchronisés. Il est bien connu que de tels effets puissent systématiquement biaiser les estimations des sensibilités des facteurs des régressions traditionnelles des séries chronologiques. Dans la littérature empirique concernant l'évaluation des actifs, plusieurs procédures de correction ont été proposées pour traiter ce phénomène pour des fréquences quotidiennes et mensuelles (*Cf.*, par exemple, Scholes et Williams, 1977 et Campbell *et alii*, 1997, pour un examen de certaines des études classiques pour cette littérature). En particulier, dans l'hypothèse simplificatrice que l'intensité d'échange et le véritable processus latent pour les rendements sont indépendants. Scholes et Williams (1977) proposent une procédure de correction qui cumule les informations additionnelles pour lisser la matrice de variance-covariance de l'échantillon. Cette procédure permet ainsi de réduire les biais microstructurels à haute fréquence. Nous complétons nos procédures d'estimation des bêtas par cette correction.

Nous montrons dans notre étude empirique que l'inclusion du troisième moment améliore sensiblement les résultats des estimations en terme de qualité du pouvoir expli-

catif des modèles et surtout de la significativité des paramètres estimés. Les gains associés à l'utilisation des données à haute fréquence, en mesurant plus exactement et en utilisant les sensibilités réalisées pour les facteurs, dépendent ainsi fortement de la qualité du modèle correspondant. Les résultats empiriques de la littérature existante suggèrent que les anomalies d'évaluation d'actifs liées à la taille et au style disparaissent en grande partie dans le modèle à trois facteurs de Fama et French (1995). Tout comme dans Fama et French (1995), le modèle à haute fréquence avec les portefeuilles classés suivant le style et la taille met en évidence une relation entre le rendement, la taille de la firme et le style de l'actif. En effet, comme mentionné par Cochrane (2001), les titres de petites entreprises avec de faibles rapports *book-to-market* font généralement de petites performances. Les volatilités réalisées moyennes au cours de la période récente se sont généralement situées dans l'intervalle de confiance des écart-types historiques de l'échantillon de longue durée. Les écart-types d'échantillon des sensibilités mensuelles suggèrent que les changements de facteurs réalisés varient également d'une manière sensible dans le temps. Les paramètres des plus grands et plus bas portefeuilles sur la variable *book-to-market* sont généralement moins volatils.

Pour accentuer la valeur économique des prévisions des modèles d'évaluation de sensibilités des actifs basés sur les données de haute fréquence, nous avons comparé la valeur *a posteriori* des portefeuilles optimaux suivant le critère moyenne-variance et des portefeuilles assortis de quelques procédures de prévision. Afin de concentrer les comparaisons directement sur l'impact des prévisions pour les sensibilités des facteurs, nous avons postulé pour cela que la matrice de covariance pour tous les portefeuilles de test, à la base des optimisations moyenne-variance, était la même pour toutes les procédures et égale à la matrice de covariance réalisée sur la période allant jusqu'à décembre 2003 (ce qui correspond à vingt-quatre mois de période de pré-échantillon à partir de janvier 2002). Nous trouvons que les modèles basés sur l'estimation des paramètres à partir des données de haute fréquence sont nettement plus prédictifs que les modèles classiques. Ce résultat est sans surprise dans la mesure où les caractéristiques intrinsèques des données sont

exploitées et incorporées dans les estimations avec des méthodologies robustes.

En définitive, suivant l'article séminal de Bollerslev et Zhang (2003) utilisant des mesures et des co-mesures « réalisées » de risque dans quelques modèles classiques d'évaluation d'actifs, ce chapitre prolonge leur approche dans un cadre multi-moment conditionnel réalisé. Ces prolongements incluent des modèles d'évaluation d'actifs d'ordre supérieur (Cf. Jurczenko et Maillet, 2006) et des modèles conditionnels d'évaluation d'actifs (Cf. Bollerslev *et alii*, 1988 et Jondeau et Rockinger, 2004). Nous avons également utilisé plusieurs méthodologies visant à neutraliser des erreurs de spécification erronées de mesures et de modèle de données (Cf. Scholes et Williams, 1977 et Ledoit et Wolf, 2003 et 2004), traitant au mieux des interdépendances entre les actifs financiers en termes de rendements (Cf. Zellner, 1962 et Parks, 1967). Des données à haute fréquence pour quelques titres sur le marché français ont été employées dans l'étude empirique relative à ce chapitre. Nous avons alors montré que les résultats des tests du CAPM étaient étroitement liés au modèle spécifié et à la méthode d'estimation utilisée. Les moments d'ordre supérieur ont néanmoins clairement un impact significatif sur la qualité des modèles. Ce résultat confirme le fait que les agents sont, non seulement sensibles aux gains espérés et aux risques potentiels, mais aussi à l'asymétrie de la distribution des rendements.

Ce chapitre montre comment à partir des transactions intra-journalières à haute fréquence, les prix des différents portefeuilles et du rendement fondamental des facteurs, on peut obtenir plus efficacement les mesures et les modèles de co-variations réalisées. Comme dans Bollerslev et Zhang (2003), nous avons expressément concentré notre analyse empirique sur un horizon des rendements d'un mois et les portefeuilles de test de Fama et French (1995). Nous constatons que, sur le marché français des actions, seul le facteur de marché a une sensibilité significative.

Au-delà de la portée de ce chapitre, il serait intéressant d'évaluer formellement l'im-

portance des erreurs de mesure des sensibilités aux facteurs dans le contexte actuel de mesures et co-mesures réalisées pour le *CAPM*. Bai *et alii* (2000), Barndorff-Nielson et Shephard (2002a, b), Meddahi (2002) ou récemment Griffin et Oomen (2008), fournissent des résultats théoriques pour le cas des mesures réalisées de volatilité. La volatilité instantanée est alors décomposée en volatilité réalisée et biais d'auto-corrélation. Griffin et Oomen (2008) montrent théoriquement et empiriquement que le biais d'auto-corrélation diminue avec la fréquence d'échantillonnage des prix pour être négligeable après 25 minutes. Un travail similaire pourrait donc être fait sur le bêta réalisé et la *skewness* réalisée. De même, en utilisant un modèle à facteurs indépendants, les différentes relations entre les actifs pourraient être simplifiées. En effet, par une extraction des facteurs communs entre les actifs par une Analyse en Composantes Indépendantes avec filtrage de bruits (*Cf.* Kopriva et Seršić, 2008), les matrices des cumulants seront toutes diagonales (du fait de la propriété d'indépendance des facteurs). De ce fait, l'estimation des paramètres avec les moments d'ordre supérieur pourra s'effectuer de manière simple avec les méthodes traditionnelles. Enfin, inspirée par les travaux de Brock et Hommes (1998) ou de Malevergne et Sornette (2006), la généralisation des approches précédentes dans un cadre hétérogène - où les agents confrontent leur fourchette de prix - pourrait constituer une extension naturelle de cette étude. Pour cela, nous pouvons envisager plusieurs sources d'hétérogénéité. En effet, nous pouvons supposer que les agents ont des points de vue différents par rapport aux anticipations des prix (Brock et Hommes, 1998). Nous pouvons aussi supposer que l'hétérogénéité provient de la perception du risque par les agents. De ce fait, plusieurs mesures de risque pour les allocations optimales entre actifs risqués et non risqués peuvent être utilisées conjointement dans un cadre multi-varié (Malevergne et Sornette, 2006).

Résumé du chapitre sur “*Recovering an Intrinsic Well-behaved Business Time from Market Prices*”

Le dernier chapitre de cette thèse se propose de trouver une fréquence d'échantillonnage des prix des sous-jacents permettant d'avoir une distribution proche de celle de la loi gaussienne. Dans la littérature empirique financière, les caractéristiques des rendements calculés en temps calendaire sont connues pour être incompatibles avec les hypothèses de Gauss-Markov, excepté à de très basses fréquences. Une façon de résoudre ce problème est d'utiliser un échantillonnage des prix à haute fréquence à différentes dates suivant les observations sur des variables exogènes. Le « temps d'échanges » est déduit des variables extraites à partir des séries de rendements comme la volatilité instantanée et la *duration*, ou des variables complémentaires comme le volume et le nombre de transactions. Il a été suggéré pour cela que pour tout processus de prix en temps calendaire, il existe un temps d'échanges où celui-ci est indépendant et gaussien.

En effet, les prix des actifs se déplacent en réponse à l'arrivée de nouvelles informations. D'une façon générale, pendant des jours, les cours peuvent avoir des variations très faibles - sans nouvelle spécifique affectant les prix - tandis que d'autres périodes peuvent être plus turbulentes avec l'arrivée de plusieurs informations. Dans le meilleur des cas, nous devons alors prêter plus d'attention au marché lorsqu'il est intense avec des mouvements rapides, que lorsqu'il est lent avec des prix qui fluctuent modérément. Formellement, Clark (1973) a présenté l'idée que la forme de la distribution des rendements des actifs financiers peut être due à une superposition des processus stochastiques, le processus des prix étant alors subordonné à un processus général d'information. Dans le cadre de cette hypothèse de mélange, les distributions des rendements d'actifs sont régies par un premier processus stochastique - un modèle classique de diffusion, lui-même conduit par une variable qui est la représentation du flux d'informations sur le marché, et modélisé comme deuxième processus stochastique. La volatilité des rendements est alors

aléatoire et considérée gouvernée par une arrivée stochastique de nouvelles informations. Celles-ci définissent alors un temps opérationnel, ou temps d'échange, différent de la mesure habituelle du temps calendaire.

Dans le temps d'échange, les prix enregistrés ne sont ainsi plus équidistants en temps réel comme la série (quotidien, hebdomadaire, mensuel, ...), mais équidistants en terme d'informations. Dans la pratique, les volumes ont été le plus souvent employés comme critères d'informations (Cf. Epps et Epps, 1976 et Harris, 1986), mais aussi le nombre de transactions ou la volatilité de la série (Cf. le *theta-time* de Dacorogna *et alii*, 2001). Ané et Geman (2000) et Geman *et alii* (2002) fournissent la forme théorique et pratique intégrant ces points, en transformant les prix en temps calendaire à partir d'un mouvement brownien standard avec des conditions raisonnables. Ils montrent avec certitude que l'intensité (ou le nombre de transactions) peut jouer un rôle majeur. Leurs résultats ont été confirmés plus tard par Gabaix *et alii* (2003), mais récemment mis en défaut par Guillemot *et alii* (2005). Ces résultats tendent néanmoins à conclure que le choix du temps calendaire n'est pas évident. En d'autres termes, n'importe quel temps d'échange est loin de mener à une série de rendements pertinents avec une distribution gaussienne et homoscedastique sans phénomène de regroupement (*clustering*) de volatilité.

Nous nous proposons dans ce chapitre de développer une stratégie pour extraire le temps implicite, de façon à lier l'activité de marché et l'arrivée de l'information pour obtenir des séries financières normalisées et propres - *i.e.* conformément aux hypothèses traditionnelles dans le paradigme moyenne-variance - pour des usages pratiques, tels que l'évaluation d'actifs et l'attribution de performances ou la gestion des risques. A la différence d'une transformation simple de type *Box-Cox*, tous les rendements de la nouvelle série normalisée correspondront aux « vrais » prix observés échantillonnés selon la circulation de l'information. Notre intuition simple est qu'il est possible d'obtenir des améliorations dans les applications financières en utilisant les véritables séries de prix avec de bonnes propriétés statistiques, simplement en changeant le couple date/heure

des échantillonnages et en évitant le caractère arbitraire du choix du prix de clôture.

Pour ce faire, nous proposons une procédure d'extraction des temps implicites à partir des données en temps calendaire seul. Cette procédure utilise les séries chronologiques des prix des actifs et délivre des séries échantillonnées à basse fréquence. Les rendements calculés pour ces fréquences tendent autant que possible vers la distribution gaussienne comme définie pour la fonction objective. L'optimisation directe classique n'est ainsi pas appropriée pour ces séries de haute fréquence parce que les paramètres à ajuster sont liés à la synchronisation de chaque échantillon et des paramètres non fonctionnels. Pour cela, les algorithmes génétiques (Holland, 1998) sont mieux adaptés, puisqu'ils peuvent réellement traiter des problèmes assez grands en taille et complexes, malgré les temps de convergence prolongés. De plus, cet algorithme qui est stochastique (plusieurs solutions qui répondent aux critères d'arrêt peuvent être sélectionnées à cause des paramètres aléatoires de départ) est conçu pour travailler avec des paramètres discrets.

En suivant l'analogie génétique, chaque solution donnée par l'algorithme s'appelle un chromosome et l'ensemble courant d'individus s'appelle le patrimoine héréditaire. Ici, chaque chromosome est un ensemble de nombres croissants représentant le nombre d'observations dans la série originale à un moment donné (jour, heure et minute). La longueur de chaque génome est le nombre d'observations désirées dans le nouveau temps d'échange de la série. La taille de la population est un compromis entre la vitesse d'exécution et l'écart minimal avec la solution voulue pour maintenir la compétitivité de l'algorithme (nous avons fixé, après plusieurs simulations sur des données réelles, la taille de la population à 300).

La fonction objective à maximiser s'appelle dans notre contexte la fonction d'ajustement, elle est utilisée pour ranger les individus dans la population, du meilleur ajustement au moins bon. Du fait des diverses manières d'examiner la normalité d'un ensemble d'observations, le choix d'une fonction objective spécifique dépend du but de l'optimisation.

Nous utilisons ici deux critères différents (la statistique de Jarque-Bera et la statistique d'Anderson-Darling) en les employant soit seuls, soit en les associant. Ce choix de Jarque et Bera et de Anderson-Darling est motivé par le fait que la statistique de Jarque-Bera incorpore les moments d'ordres trois (Skewness) et quatre (*Excess kurtosis*) dans sa définition. La statistique d'Anderson-Darling permet, quant à elle, de tester le meilleur ajustement en utilisant une distribution logarithmique. De ce fait, elle accorde plus d'importance aux queues de distribution. Sa distribution dans le cas gaussien étant tabulée.

Comme la plupart des algorithmes d'optimisation numérique stochastique, les algorithmes génétiques travaillent de façon itérative, en générant des solutions meilleures à chaque itération jusqu'à ce que les conditions d'arrêt soient remplies. A partir d'une population donnée, une itération crée une nouvelle population de même taille appelée génération en appliquant deux opérateurs. Des opérateurs génétiques sont appliqués à la population de parents pour créer le descendant. En conséquence, la première étape est alors le choix des parents, ensuite vient le choix des opérateurs. Cinq procédures principales existent pour choisir les parents : le meilleur choix (*Best method*), la méthode de tournoi (*Tournament method*), la méthode d'ajustement (*Fitness method*), la méthode d'ajustement par sur-sélection (*Fitness-overselection method*) et la méthode aléatoire (*Random method*). Dans chaque cas, la méthode doit permettre d'atteindre les deux objectifs : sélectionner les meilleurs individus (pour améliorer la fonction objective) et maintenir la diversité dans la population (pour éviter l'attraction des *extremum* locaux). Bien que les résultats soient à peu près semblables quelle que soit la méthode choisie, Chidambaran *et alii* (1999) trouvent des différences significatives dans la vitesse de convergence suivant les approches. Le meilleur algorithme doit donc pouvoir converger assez rapidement sans toutefois perdre en qualité d'information.

Dans la méthode aléatoire, des individus parents sont choisis indépendamment de leurs points d'ajustement. De ce fait, elle fournit une dérive aléatoire pure à travers l'espace des solutions possibles. Dans la méthode d'ajustement, les populations d'indi-

vidus de la prochaine génération sont choisies toujours aléatoirement, mais pondérées par leur coefficient d'ajustement. Dans la méthode de tournoi, les meilleurs individus dans un échantillon aléatoire sont gardés, tandis que dans la méthode d'ajustement par sur-sélection, les individus sont regroupés selon leurs points d'ajustement avant d'être choisis avec une probabilité différente pour chaque groupe. Nous avons utilisé dans nos estimations, la méthode du meilleur choix dans laquelle les meilleurs individus, rangés par degré d'ajustement, sont utilisés pour produire la prochaine génération.

Comme pour le choix des populations de parents, il existe plusieurs procédures pour l'opérateur de reproduction. Le but est de multiplier les nouveaux individus de la population choisie parmi les meilleures solutions. La manière la plus simple de procéder serait de reproduire simplement la nouvelle population dans les génomes des parents. Cependant, cette solution se fonde seulement sur la mutation des individus pour apporter de nouvelles solutions dans le patrimoine héréditaire, et est par conséquent relativement inefficace. Une meilleure manière de procéder est de combiner les caractéristiques des parents dans un processus analogue à la reproduction sexuée. Cette technique, connue sous le nom de croisement, consiste à prendre le génome d'un des parents, et de l'échanger avec les gènes correspondants de l'autre parent. Dans le croisement simple que nous utilisons dans ce chapitre, un pivot simple est choisi aléatoirement parmi les gènes de l'échantillon. Les deux gènes des parents sont comparés et un nouvel individu est créé. Les rôles des premier et deuxième parents sont permutés pendant le temps nécessaire pour assurer que l'enfant soit un individu viable.

Pour que l'algorithme converge, il faut que plusieurs solutions remplissent les conditions d'arrêt, *i.e.* que la statistique de normalité d'Anderson-Darling ne permette pas de rejeter l'hypothèse nulle d'une distribution gaussienne du temps d'échange. Bien que Ané et Geman (2000) avancent qu'il existe un temps de transactions où le processus de prix des actifs est un mouvement brownien, il apparaît moins évident dans la pratique. Au fur et à mesure que la fréquence augmente, il est clair qu'il deviendra de plus en plus difficile

de trouver ce temps de transactions. En effet, l'écart par rapport à la normalité est plus élevé d'une part, et d'autre part, les possibilités de prélèvements possibles sont moins nombreuses. Bien que nous présentions ici le résultat pour des données quotidiennes, les expériences préliminaires suggèrent que l'existence d'une échelle de temps de transactions pour de plus hautes fréquences est difficile. La convergence peut être accélérée par plusieurs choix pratiques, notamment un choix instruit des paramètres de départ. Nous utilisons des paramètres de départ totalement aléatoires, mais explorons aussi plusieurs autres conditions initiales pour essayer d'améliorer le processus. Dans le cas des données quotidiennes, il pourrait y avoir plusieurs solutions respectant les contraintes, et la théorie ne donne pas un critère de choix spécifique.

Les données utilisées pour tester notre procédure d'extraction proviennent d'*Euronext* et comportent les prix de 41 titres, entre le 02 janvier 2002 et le 29 décembre 2006, prélevés à intervalle de 5 minutes. Ainsi, toutes les cinq minutes, le dernier prix de chaque action est enregistré sur *Euronext*. Les périodes où tous les titres n'étaient pas renseignés ont été éliminées de la base. L'échantillon comporte au total 1281 jours ; ce qui fait en moyenne 1,4 jour calendaire entre chaque prix de clôture (quand on inclut les week-ends, les vacances et les jours fériés). Le nombre d'observations par jour n'est ainsi pas parfaitement cylindré, d'autant plus que les heures d'ouverture d'*Euronext* ont changé au cours de la période considérée (prolongées d'une demi-heure). Les prix de clôture quotidiens étant les observations les plus communément disponibles, nous avons décidé ici de les employer comme repère, bien que nous ayons expérimenté d'autres fréquences. Chaque gène correspond à un nombre entier ; la position de l'observation dans la série originale échantillonnée à 5 minutes (qui correspond à des dates pour lesquelles un prix est disponible). Le génome fournit ainsi finalement une série de prix, dont les rendements seront utilisés comme une entrée de la fonction objective. Ainsi, le temps de transaction, pour cette sous base de référence comporte au final le même nombre d'observations que le nombre de jours de l'échantillon. Toutes les quantités sont ainsi basées sur le même ordre de grandeur que le nombre de jour des données, et le nombre moyen d'observations par

jour de marché est par construction égal à un.

Il est à noter que le processus d'optimisation, qui a été répété sur les 41 titres selon le critère d'arrêt d'Anderson-Darling comme expliqué ci-dessus, converge toujours sur notre échantillon. Ainsi, cette convergence assez régulière de l'algorithme d'optimisation nous fournit un pas d'échantillonnage pour lequel la distribution des rendements des différents actifs est gaussienne. Comme remarques particulières sur les séries extraites, on observe de façon nette les effets de la déformation du temps. En effet, les queues de distribution des séries extraites sont aplaties avec une diminution du taux d'apparition des valeurs extrêmes. Nous trouvons aussi que l'algorithme d'échantillonnage ne conserve pas les prix qui diffusent peu d'informations (notamment les prix de mai 2005 où les marchés étaient très calmes). Il convient de noter qu'à la différence d'un taux de change qui est continu, la base de séries chronologiques pour les actions est non continue à cause des biais des périodes de fermeture (pendant la nuit par exemple) ; de ce fait, il y a des sauts irréductibles dans les prix entre l'heure exacte de fermeture et le moment suivant l'ouverture, ce qui explique que quelques rendements extrêmes demeurent dans la base extraite (puisqu'ils ne peuvent pas être divisés en périodes plus courtes).

Bien que le critère de normalité soit dans ce cas-ci le test d'ajustement d'Anderson-Darling, nous trouvons que la déformation de temps réduit considérablement l'amplitude de la *skewness* et de la *kurtosis*, alors que les rendements en temps calendaire montrent une déformation classique sur les queues de distribution (queues épaisses, asymétrie élevée, ...). Les tests de normalité appliqués sur les séries ainsi extraites confirment ce résultat. Les temps de transaction sont en outre extrêmement sensibles aux observations produisant de grands mouvements de marché ce qui entraîne une accumulation d'enregistrements d'informations les jours de fortes volatilités. Toutefois, il n'y a pas une indication claire des sens de variation sur le régime des prix produit par le processus. En effet, tandis que les jours pour lesquels plusieurs observations sont prises en compte tendent à se reproduire (été 2003 par exemple), la persistance n'est pas extrêmement marquée, et

de plus il n'y a pas une division de l'échantillon entre des périodes de fortes volatilités et les périodes de faibles volatilités (le taux de prélèvement demeure donc relativement uniforme pour l'échantillon entier).

En analysant la duration (réciproque de la fréquence) définie comme l'intervalle de temps entre deux observations, on constate qu'elle est étalée sur une large bande. En effet, la distribution de la série de duration fournit un minimum de 5 minutes (qui correspond au pas d'échantillonnage de la base de données) et un maximum de 78 jours (ce qui correspond à peu près à 3 mois sans prise en compte d'une information sur le prix du titre). La distribution des temps de transaction est ainsi fortement asymétrique, avec un pic à un jour. Les discontinuités observées dans la fonction correspondent aux intervalles pour lesquels les données n'existent pas (principalement l'existence d'une période durant la nuit). Ce résultat confirme donc le fait que les prix des actifs ne suivent pas forcément une loi de distribution unique, ceci à cause des multiples facteurs externes qui inter-agissent de façon complémentaire et des niveaux différents pour modifier la structure des prix des actifs sur les marchés.

De même, en comparant, en premier lieu par une extraction action par action, les corrélations des rendements en temps calendaire et en temps de transactions, on remarque que ces derniers sont très faibles pour chaque action. Ce résultat est principalement dû au fait que les séries extraites ne sont pas rangées suivant le même pas ; ce qui entraîne une destruction des inter-relations entre les séries. De plus, aucune des corrélations entre actifs en temps de transactions n'est significative. Ces résultats négatifs pour l'estimation actions par actions nous ont fait recourir en deuxième lieu à une procédure d'extraction globale avec un même pas d'échantillonnage afin de choisir des temps de transactions synchronisés pour tous les titres de l'étude (même si nous pouvons nous attendre dans ce cas à une perte de la qualité des ajustements à la loi normale).

En estimant simultanément les temps de transactions de tous les titres de l'étude,

nous trouvons une forte augmentation des valeurs de la matrice de corrélations entre les rendements en temps calendaire et en temps de transactions. Ce résultat confirme bien le fait qu'il faut avoir une approche globale pour les estimations des problématiques de synchronisation et comparabilité des différents titres. Enfin, pour tester l'efficacité de notre méthode d'extraction de temps, nous avons construit à partir des matrices de corrélations, les portefeuilles de variance minimale. Les résultats sont plutôt mitigés. En effet, les portefeuilles déterminés à partir des estimations sur le temps calendaire et ceux estimés en temps de transaction rapportent un rendement moyen annualisé absolu d'environ 8,5%.

En définitive, nous avons dans ce chapitre soulevé et répondu à deux problématiques fondamentales en ce qui concerne les données de très haute fréquence. D'une part, nous avons présenté un cadre pratique pour extraire les temps implicites pour le ré-échantillonnage des séries financières réelles et d'autre part, nous avons montré que les résultats des séries transformées sur les principaux modèles financiers classiques étaient plus robustes; ceci par l'élimination du biais de l'hypothèse de normalité des rendements. La réponse à ces problématiques a été possible grâce à l'exploitation efficiente de la technique de l'algorithme génétique avec comme critère de convergence de la fonction objective la gaussianité de la distribution des rendements. Néanmoins, certains problèmes demeurent dans l'utilisation de cet algorithme, notamment la dépendance au choix de la mesure exacte à utiliser ou bien la non-unicité des solutions optimales. Nous aurions pu aussi estimer simultanément les rendements des titres étudiés par un critère multi-varié de normalité. De même, nous avons extrait les séries de rendements ayant une distribution gaussienne (par hypothèse de la fonction objective de l'optimisation); d'autres lois de distribution peuvent être testées et utilisées.

Cette nouvelle série en temps intrinsèque avec des propriétés statistiques standard, ouvre une large voie pour les recherches pratiques ultérieures. Notamment, les futures études sur l'évaluation d'actifs, l'attribution de performances ou bien la gestion des risques

pourraient s'appuyer sur des données implicites pour les estimations empiriques. Dans ce cas de figure, l'invariance recherchée dans la distribution des rendements est le résultat d'une compréhension globale des fluctuations des prix et des fluctuations du temps. Le temps boursier n'est alors plus indépendant de l'action des opérateurs, mais étroitement lié aux variables comme la taille du marché, le volume de l'échange et le nombre de participants de la bourse. L'activité des opérateurs agit comme une pondération des durées, en contractant plus ou moins l'espace des cotations, en déformant le temps calendaire physique exogène du marché, en temps intrinsèque endogène au marché. Il est possible d'envisager un temps implicite pour chacune de ces variables et une agrégation multi-variée de ces différents temps pour une meilleure prise en compte des vraies sources de variations des prix des actifs. De ce fait, la valeur d'un prix coté (et donc la pertinence de la mesure) n'est pas la même selon que ce prix correspond à une transaction de cinq cent mille titres ou de cinq titres. Cette intuition qui trouve une traduction scientifique dans les modèles à temps déformé (Salmon et Tham, 2007) traduit le fait que sur les marchés financiers, la durée du temps n'est pas égale à la valeur du temps. Toutefois, le temps implicite déterminé par les algorithmes proposés dans ce chapitre ne fait que transférer le problème de l'hypothèse distributionnelle en temps calendaire au problème de prévision des prix et correspondance des valeurs en temps déformé. Néanmoins, tout comme pour les modèles à duration (Engle et Russell, 1997), où les informations sur les prix ne sont pas échantillonnées sur un intervalle de temps régulier, le temps implicite pourrait servir de base à de nouveaux modèles sur un espace particulier (espace du volume des échanges, du nombre de participants ou bien du niveau de variation du prix). De plus, de nouveaux produits financiers pour des agents particuliers, exprimés en temps implicite, pourraient voir le jour. Ces produits pourront alors être définis suivant des critères exprimés en temps déformé (volatilité constante sur cet espace, valeurs liquidatives enregistrées suivant le volume atteint, ...).

Chapitre 1

« Roses des vents », « éventails » et « explosions d'étoiles » sur le marché français : caractérisations, mesures et applications

1.1 Introduction

L'impact des phénomènes de microstructures des marchés sur le processus de formation des prix des actifs a depuis des décennies fait l'objet de recherches aussi bien théoriques qu'empiriques. Ces recherches ont montré que les composantes de la fourchette de prix affichée (*Bid-Ask*) dépendaient principalement de trois facteurs : le coût de traitement des ordres (*order processing cost*, Demsetz, 1968) ; le coût d'opportunité d'une position mal diversifiée (*inventory holding cost*, Ho et Stoll, 1983 ; Biais, 1993) et le coût de sélection adverse (*adverse information cost*, Glosten et Milgrom, 1985). Mais ces composantes théoriques ne permettent pas d'expliquer certains phénomènes comme l'existence de seuils ou de barrières psychologiques sur les marchés boursiers (De Grauwe et Decupere, 1992 et Morel et Teïletche, 2001) ou bien encore la concentration des prix autour

de certaines valeurs (Gottlieb et Kalay, 1985 ; Harris, 1990 et 1991 ; Anshuman et Kalay, 1998 ; Huang et Stoll, 1994 et 2001 et Capelle-Blancard et Chaudhury, 2007).

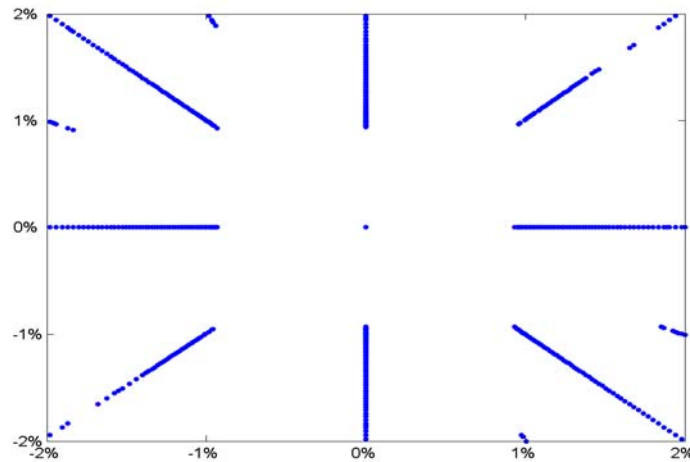
Ces seuils et barrières pouvant être à l'origine de la modification des processus générateurs de la structure des prix, une littérature récente, originellement inspirée par les études sur la théorie du chaos, est consacrée à la détermination de la structure des rendements des titres sur les marchés financiers. Ces études, dans leurs versions les plus simples, consistent à construire un diagramme de phases des rendements ou *Scatter Plot* en anglais (représentation graphique des x_t contre les x_{t+1}) et d'interpréter les résultats obtenus. Il ressort de ces études, une distribution non uniforme des points représentés dans le diagramme de phases indiquant la présence d'une structure complexe. Cette structure a la particularité de pouvoir affecter significativement les paramètres des différents processus utilisés pour modéliser les rendements des titres (Ball, 1988, Annaert *et alii*, 2003). En effet, les prix des titres, déterminés par les variations minimales de cours ou échelon de cotation, ne sont pas par essence, des variables continues mais discrètes. Cette caractéristique restreint les rendements des titres à prendre un nombre limité de valeurs. Cette restriction est une condition nécessaire à l'apparition de ce qui est appelé une « rose des vents »¹ (ou *Compass Rose* en anglais, voir graphique 1.1 présentant le diagramme de phases des rendements d'un titre). Cette structure prend la forme de droites espacées, épaisses, émergeant de l'origine et fuyant vers les principales directions des points cardinaux, d'où le nom de « rose des vents » emprunté aux navigateurs.

Crack et Ledoit (1996) ont été les premiers à découvrir une telle représentation graphique sur les rendements journaliers des titres du *New York Stock Exchange* (NYSE) et ils décrivent le phénomène de rose des vents comme étant subjectif. L'existence de cette structure, qui est un artefact de la microstructure des marchés financiers², ouvre plusieurs

¹Une rose des vents est la partie mobile d'un compas servant de boussole aux navigateurs. Elle indique les différents points cardinaux et inter-cardinaux.

²Depuis le milieu des années 80, plusieurs articles - parmi lesquels ceux de Holthausen *et alii* (1987), Aitken et Frino (1996), Chan et Lakonishok (1993, 1995, 1997), Keim et Madhavan (1995, 1996, 1997) et

FIG. 1.1. Diagramme de phases des rentabilités



Source : simulation de 30 000 points à partir d'un modèle $AR(2) - GARCH(1,1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1,1)$ sont respectivement de $\{5, 2 \cdot 10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2 \cdot 10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondante est ensuite déterminée par discrétisation des prix en suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR et un pas d'échelon de cotation de 1 EUR. L'axe horizontal représente les rendements du titre à une période donnée; l'axe vertical représente les rendements à la période précédente. Les axes sont exprimés en %.

voies de recherches, notamment sur les déterminants de son apparition, sur sa quantifica-

Gallo (2001) - ont montré qu'il existe une forme particulière de biais microstructurel, liée à la « danse de la fourchette » de prix (*bid-ask bounce* en anglais), qui affecte la structure auto-corrélative des rendements des actifs. Ce biais proviendrait de l'asymétrie qui existe entre les blocs vendeurs et les blocs acheteurs (les blocs acheteurs ayant tendance à faire baisser les prix et les blocs vendeurs à les augmenter). La structure auto-corrélative des rendements sera ainsi modifiée par ces deux effets successifs : un prix en hausse sera en général immédiatement suivi d'un prix en baisse et *vice-versa* (Frino *et alii*, 2004), ce qui laissera une trace dans l'auto-corrélogramme. L'auto-corrélation d'ordre 1 résultante aura ainsi une valeur négative du fait de ce biais. Mais le phénomène de « rose des vents » analysé ici est cependant plus complexe que celui de la « danse de la fourchette », car il intègre aussi d'autres formes de biais microstructurels. Ainsi, en étudiant directement les variations de prix, nous prenons en compte l'ensemble des facteurs qui peuvent les influencer. Notons ici que le phénomène de rose des vents, qui est principalement une conséquence de l'imposition d'un pas de variation des prix d'un titre, pourrait être amplifié par un effet systématique du type « danse de la fourchette », et par les actions des agents ayant une attirance pour certains chiffres (Morel et Teïletche, 2001). Ces effets se matérialiseront par une déformation de la « vraie » structure d'auto-corrélation des rendements de l'actif.

tion et sur les biais potentiels qu'elle peut causer dans les estimations des modèles. Szpiro (1998) par exemple montre que ce phénomène apparaît sur des titres individuels aussi bien sur des portefeuilles de plus de deux actifs sur le marché des changes. La forme d'une rose des vents cacherait en fait une structure plus complexe qu'il nomme « nano-structure ». Selon son étude, seul le changement discret de prix est une condition nécessaire et suffisante à l'apparition d'une rose des vents. Cependant, cette discrétisation des prix n'a pas été examinée en détail, alors que la persistance d'un phénomène de rose des vents peut avoir des implications importantes. En effet, selon Amilon et Byström (2000), Koppl et Tuluca (2004), Amilon (2003) et Krämer et Runde (1997), la discrétisation a l'effet d'induire des erreurs d'arrondis sur les prix observés par rapport aux « vraies » valeurs instantanées des titres. Gottlieb et Kalay (1985) montrent par exemple que la discrétisation des prix constitue une source de biais dans les estimations des variances des rendements des titres et sur les paramètres des estimations des principaux modèles économétriques appliqués directement sur des données de haute fréquence. Dans ce contexte, il s'avère nécessaire de trouver une caractérisation propre à ce phénomène pour mieux le cerner.

A ce propos, plusieurs manières de rendre compte du phénomène ont déjà été recensées dans la littérature. Une première idée de caractérisation, développée par Wang et Wang (2002) et corrigée par Mitchell et McKenzie (2006), est basée sur l'analyse de l'alignement des droites passant par les différents points des différentes directions d'une rose des vents parfaite (couplets de rendements successifs) du *Scatter Plot* (ou diagramme de phases) et rejoignant l'origine. Plus l'on trouvera des droites alignées dans les bonnes directions (Nord, Sud, Est, Ouest, ...) et plus le phénomène sera caractérisé. Une seconde idée, développée à l'origine par Koppl et Nardone (2001), très similaire à la précédente, repose sur la distribution des angles de ces mêmes droites dans le diagramme de phases. Si les points se regroupent en un nombre limité d'endroits dans le *Scatter Plot*, alors les angles, formés par les droites passant par ces points et rejoignant l'origine, ne prendront qu'un nombre limité de valeurs et nous devrions en trouver trace dans la densité des coordonnées polaires des points du *Scatter Plot*. La troisième idée dont nous nous ser-

virens repose sur le constat que dans le cas d'un phénomène marqué de rose des vents, seules quelques parties du *Scatter Plot* seront remplies de points, les autres étant vides. En représentant le diagramme de phases sous forme d'un tableau (de contingence) issu de la discrétisation des axes, il devient alors possible de caractériser le phénomène par l'importance relative des parties vides du tableau (pixels blancs) par rapport aux cases remplies de couplets de rendements successifs (pixels noirs).

Ces trois idées de base sous-tendent la définition des trois indicateurs d'intensité du phénomène de *Compass Rose* ($CRII_1$, $CRII_2$ et $CRII_3$) - mesures complémentaires et originales fondées sur plusieurs définitions du phénomène - que nous proposons dans la suite de cet article afin de caractériser plus avant le phénomène de « rose des vents ». Grâce à ces mesures, nous confirmons tout d'abord, sur la base de plusieurs types de simulations réalistes, que ce phénomène résulte bien principalement d'un effet de discrétisation (plus ou moins sensible en fonction de la volatilité), de la période d'échantillonnage, mais aussi de la taille de l'échantillon étudié. Enfin, les mesures que nous effectuons sur quelques titres du marché français confirment aussi l'hypothèse que ce phénomène est assez général et à ce titre, digne d'intérêt.

Néanmoins, comme l'écrivent Mitchell et McKenzie (2006, page 3 519), le phénomène ne revêt pas *a priori* « d'intérêt particulier pratique » direct. Toutefois, l'étude de la « rose des vents » présentée dans cet article, au-delà de son aspect de curiosité et de son caractère phénoménologique qui nous amène à nous intéresser aujourd'hui de plus en plus aux données de haute fréquence, permet aussi de s'intéresser indirectement à certains autres aspects pratiques liés au choix de la fréquence d'observation pertinente. Ainsi, le choix des seuils d'agrégation temporelle acceptables pour l'utilisation de modèles financiers sur des données de haute fréquence pourrait-il être guidé aussi par l'absence de phénomène³. En effet, un courant récent de la littérature s'intéresse au contenu informa-

³Notons de plus que, comme dans le cas de la « danse de la fourchette » (auto-corrélation d'ordre 1 négative en haute fréquence, voir graphique A.3 en annexe), et contrairement au phénomène des seuils psychologiques, aucun article n'a jusqu'à présent, à notre connaissance, mis en avant une possible

tionnel des données de haute fréquence (Cf. Goodhart et O'Hara, 1997 et Woehrmann, 2007). Des grandeurs dites « réalisées », calculées à partir de données de haute fréquence, ont été proposées récemment comme la volatilité (Cf. Andersen *et alii*, 2000 et 2001), les *betas* (Cf. Bollerslev et Zhang, 2003), ou encore la *Value-at-Risk* (Cf. Dacorogna *et alii*, 2001 ; Giot et Laurent, 2004 et Colletaz *et alii*, 2007) pour la mesure et le contrôle des risques des *traders*. Or, à très haute fréquence (données de transactions), il a été montré qu'il existait des biais microstructurels importants dans les mesures (Cf. Oomen, 2002 et Griffin et Oomen, 2008). Ainsi, il existe donc un compromis à trouver entre, d'une part, se rapprocher le plus possible du « vrai » processus de prix des marchés en utilisant des données de haute fréquence, et d'autre part, incorporer du bruit de microstructure dans le calcul d'un estimateur théoriquement plus précis (Cf. Aït-Sahalia *et alii*, 2005 et Nielsen et Frederiksen, 2008). Il est donc intéressant de voir à partir de quel seuil d'agrégation temporelle le phénomène - très présent sur le marché français en temps de transactions - disparaît, et de définir ainsi une fréquence-type d'observation qui permette de réduire ce phénomène de bruit (Cf. Oomen, 2002).

La suite de l'article est organisée de la manière suivante. Nous commençons par décrire dans la deuxième section le phénomène de rose des vents, ses déterminants, les aspects théoriques sous-jacents et la forme théorique des mesures proposées. La troisième section sera consacrée à l'aspect empirique de l'étude. Nous indiquerons les variables utilisées, ainsi que les modèles et les méthodes employés pour nos simulations. Nous approfondirons l'étude de Gleason *et alii* (2000) et déterminerons le seuil à partir duquel une rose des vents apparaît. La dernière section sera consacrée à la conclusion et aux possibles extensions de l'étude.

exploitation du phénomène de rose des vents à des fins d'arbitrage ou de spéculation.

1.2 Définition, caractérisation et mesures d'une rose des vents

Dans la littérature financière, plusieurs définitions complémentaires pour la « rose des vents » sont proposées. Celles-ci sont étroitement liées à la visualisation du diagramme de phases. Nous présentons tout d'abord le phénomène avant de fournir les formes théoriques des mesures quantitatives qui intégreront les caractéristiques trouvées.

1.2.1 Définition et caractérisation du phénomène

Pour mieux comprendre cette étrange structure, il est nécessaire de distinguer d'une part, une définition formelle à partir de la terminologie utilisée, et d'autre part, les caractéristiques du phénomène qui peuvent être liées aux effets conjugués de la discrétisation des prix et de la corrélation sérielle des rendements. Après avoir donné une définition d'une rose des vents selon la terminologie propre à ce contexte, nous présenterons les déterminants du phénomène sur la base des différents articles dédiés à cette structure de dépendance.

1.2.1.1 Définition d'une rose des vents et de ses variantes

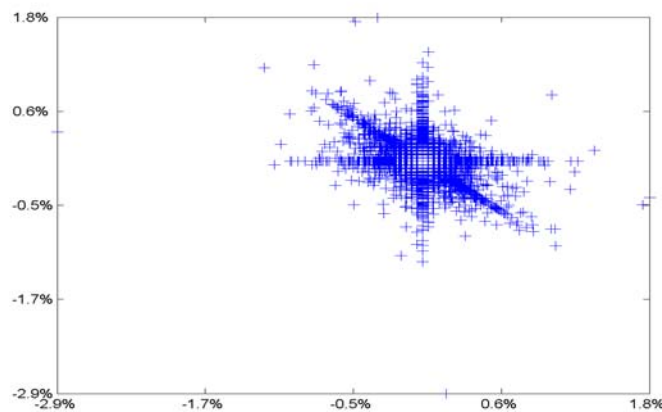
La rose des vents peut être analysée ou visualisée sous plusieurs angles ; dans un plan (structure classique ou éventail) ou dans un espace (explosion d'étoiles). Nous adoptons donc ici une démarche similaire à celle de Crack et Ledoit (1996) afin de mettre en évidence les phénomènes de rose des vents, d'éventail de volatilité et d'explosion d'étoiles.

Rose des vents (Compass Rose)

Quand on représente les rendements (intra-journaliers) à la période t , sur l'axe horizontal, contre les rendements à la période $t + 1$, sur l'axe vertical en données intra-journalières, comme le suggèrent Huang et Stoll (1994), il semble clair qu'il existe peu de relations

entre ces mouvements de prix. Crack et Ledoit (1996) ont montré que le simple choix typographique de la croix (+) symbolisant les points et le choix d'une série d'observations à court terme pouvait obscurcir une structure pourtant significative avec d'autres choix typographiques. En remplaçant les points par des croix et dès lors que l'on réduit un peu la période d'observation, la structure non-visible du graphique devient claire (voir par exemple la différence entre le graphique 1.2 et le graphique 1.3 suivant). Ce changement est marqué par des points plus étalés (« *smudge* », Szpiro, 1998), par des lignes fuyant de l'origine, par des droites épaisses et bien orientées sur les directions principales d'une boussole, d'où le nom de « rose des vents ».

FIG. 1.2. Diagramme de phases des rentabilités du titre Peugeot

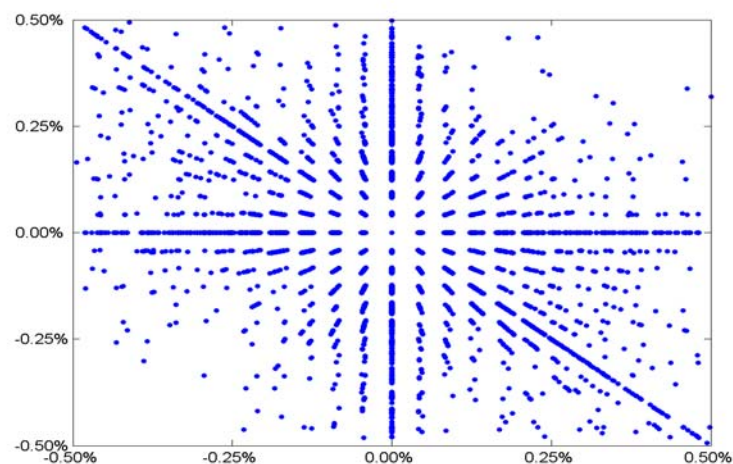


Source : Euronext. L'axe horizontal représente les rendements du titre Peugeot à une période donnée ; l'axe vertical représente les rendements à la période précédente. Le signe plus (+) est utilisé pour symboliser les points. La période d'observation s'étend du 02 janvier au 13 août 2004 (1 283 125 points). Les axes sont exprimés en %.

Eventail (Fan-shaped Structure)

En considérant le carré des rentabilités comme un *proxy* de la volatilité instantanée (Cf. Chen, 1997), il est possible de mettre en évidence une structure particulière dite « d'éventail » comme illustrée sur le graphique suivant. Ce phénomène est directement

FIG. 1.3. Diagramme de phases des rentabilités du titre Peugeot avec modification de l'échelle



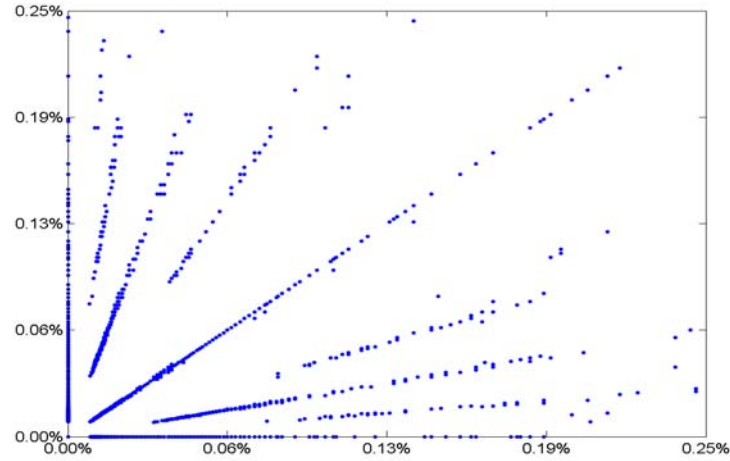
Source : Euronext. Ce graphique est similaire au graphique 1.2; seule l'échelle a été réduite et le choix typographique est différent. L'axe horizontal représente les rendements intra-journaliers à une période donnée. L'axe vertical représente les rendements intra-journaliers de la période précédente. Le symbole point (.) représente les données qui couvrent la période d'observation qui s'étend du 02 janvier au 13 août 2004 (1 283 125 points). Les axes sont exprimés en %.

lié à celui de la rose des vents, puisque lorsqu'une rose des vents est parfaitement définie, le repliement des quatre quadrants des couples de rendements laisse apparaître une sorte d'éventail.

Explosion d'étoiles (Star Burst)

Comme mentionné par Crack et Ledoit (1996), le phénomène peut avoir une dimension de plongement supérieure à 2. Ainsi dans l'espace défini par trois rendements successifs, il est possible de voir apparaître une structure particulière - appelée « explosion d'étoiles » (*star burst* en anglais) - qui traduit un phénomène d'alignement des triplets de rendements le long de droites fuyant à partir de l'origine comme illustré sur le graphique suivant.

FIG. 1.4. Phénomène de rose des vents pour les rentabilités au carré (illustration d'un éventail)



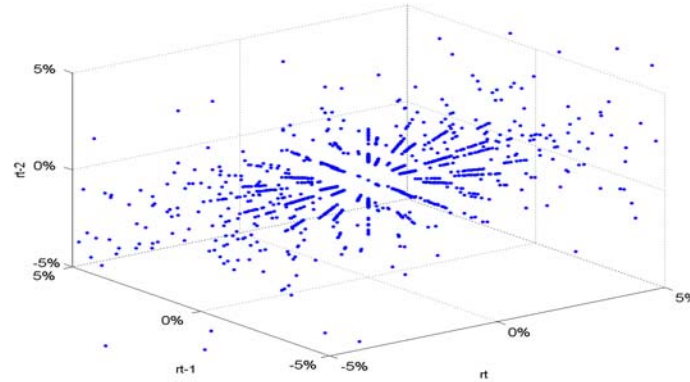
Source : Simulation de 30 000 points à partir d'un modèle $AR(2) - GARCH(1,1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1,1)$ sont respectivement de $\{5, 2.10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2.10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondant est ensuite déterminée par discrétisation des prix en suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR et un pas d'échelon de cotation de 1 EUR. L'axe horizontal représente les rendements au carré du titre à une période donnée; l'axe vertical représente les rendements au carré à la période précédente. Les axes sont exprimés en %.

Sans perte de généralité, nous étudierons dans la suite le phénomène de rose des vents, sachant que notre analyse se généralise très facilement aux phénomènes d'éventail et d'explosion d'étoiles.

1.2.1.2 Caractérisation du phénomène

Crack et Ledoit (1996) ont été les premiers à découvrir une telle représentation sur les rendements journaliers. Pour eux, le déterminant le plus important pour l'existence d'une rose des vents est la taille effective de l'échelon de cotation. En effet, ils montrent que

FIG. 1.5. Phénomène de rose des vents pour des triplets de rentabilités successives (explosion d'étoiles)



Source : simulation de 30 000 points à partir d'un modèle $AR(2) - GARCH(1,1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1,1)$ sont respectivement de $\{5, 2.10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2.10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondante est ensuite déterminée par discrétisation des prix en suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR et un pas d'échelon de cotation de 1 EUR. L'axe horizontal représente les rendements du titre à une période t donnée; l'axe horizontal représente les rendements du titre à une période $t - 1$ donnée; l'axe vertical représente les rendements à la période $t - 2$. Les axes sont exprimés en %.

la taille « officielle »⁴ de l'échelon de cotation (imposée par la bourse) n'est ni nécessaire ni suffisante pour l'apparition du phénomène. Il faut comparer la taille « effective » de l'échelon de cotation (déterminé par les participants du marché⁵) aux changements de prix. Ils trouvent aussi que l'existence d'une telle structure est liée à des sauts discrets des prix sur l'ensemble des titres échangés sur le *New York Stock Exchange* (NYSE). Lee *et alii* (1999) dans une étude empirique sur le marché à terme, arrivent aux mêmes conclusions quant aux déterminants, mais pas sur le caractère systématique du phénomène. En effet, leur étude montre que la rose des vents survient sur certains contrats à terme,

⁴La notion de taille « officielle » fait ici référence au pas minimum de variation du prix de l'actif (*tick* en anglais).

⁵Cette taille dite « effective » de l'échelon de cotation est principalement liée au degré de liquidité de l'actif échangé, contrairement à la taille réelle qui est, elle, fixe.

suggérant ainsi que la taille de l'échelon de cotation (*tick*), différente selon les contrats, peut être un facteur déterminant pour l'existence d'une rose des vents. Par contre, ils démontrent que la présence du phénomène n'est pas effective sur certaines catégories de contrats. Par exemple, la structure est totalement absente pour les titres de l'industrie et les contrats monétaires, et dominante sur les contrats sur les métaux et l'énergie. Ces résultats suggèrent l'existence d'une relation entre les catégories sectorielles et l'apparition d'une rose des vents. En revanche, pour Szpiro (1998), ce phénomène apparaît sur tous les titres individuels et aussi sur certains portefeuilles de plus de deux actifs. Sa forme cache en fait une structure plus complexe (nano-structure). Pour lui, seul le changement discret de prix est une condition nécessaire et suffisante à l'apparition d'une rose des vents. Bien que la discrétisation des prix soit avancée comme étant la condition déterminante pour l'apparition du phénomène, des études récentes (voir par exemple Gleason *et alii*, 2000) montrent que la volatilité des rendements est aussi un facteur essentiel. Plus particulièrement, le rapport entre la taille de l'échelon de cotation et la volatilité des rendements doit être au-dessus d'un certain seuil afin qu'une structure de rose des vents puisse apparaître. Les résultats de l'étude menée par Gleason *et alii* (2000) montrent que pour un écart-type de 0,02%, le seuil du ratio taille de l'échelon de cotation sur volatilité est très petit. Une structure nette apparaît lorsque le ratio est de $1/50$. Avec un écart-type de 0,6%, le seuil est de $1/4$. Ceci car la volatilité des données intrajournalières est plus faible que celle calculée sur des données journalières. Les résultats de leurs simulations expliquent pourquoi une rose des vents apparaît sur des séries intrajournalières et rarement sur celles journalières. Ils montrent donc que le ratio taille de l'échelon de cotation sur la volatilité est déterminant pour l'apparition du phénomène. Wang *et alii* (2000) illustrent l'effet de l'échelle sur la présence d'une rose des vents. Pour eux, la probabilité d'apparition du phénomène augmente avec l'échelle : plus la fréquence d'observation est haute et plus l'existence d'une telle structure est probable.

Nous pouvons donc distinguer deux catégories d'effets : un effet « discrétisation » qui est causé par la présence d'un pas d'échelon de cotation important (Crack et Ledoit, 1996 ;

Lee *et alii*, 1999 et Szpiro, 1998) et un effet « corrélation » qui provient de la dépendance temporelle qui augmente la probabilité d'apparition du phénomène de rose des vents (Gleason *et alii*, 2000 et Wang *et alii*, 2000).

1.2.2 De la mesure du phénomène à la construction des indices d'intensité du phénomène de rose des vents (*Compass Rose Intensity Indexes*)

Depuis l'article de Crack et Ledoit (1996), les travaux sur le phénomène de rose des vents sur les marchés financiers se sont attelés à déterminer les conditions d'existence et ses effets sur les tests statistiques et les prévisions. Si ce phénomène a un impact sur les principaux tests statistiques, comme le montrent Amilon et Byström (2000) ou bien Krämer et Runde (1997), on peut se demander comment le quantifier ou l'évaluer ; ceci d'autant plus qu'une analyse fondée sur une simple visualisation d'un diagramme de phases n'est pas possible pour certaines tailles d'échantillon trop grandes. Nous proposons la caractérisation du phénomène à l'aide de trois indicateurs d'intensité que nous présentons dans les prochaines sous-sections.

1.2.2.1 Un premier indice d'intensité du phénomène de rose des vents ($CRII_1$)

L'indicateur $CRII_1$ est basé sur l'idée suivante (Wang et Wang, 2002 et Mitchell et Mckenzie, 2006) : la « rose des vents » est bien « définie » (*i.e.* parfaitement visible et à l'image d'un compas marin classique) lorsque toutes les droites la représentant (*i.e.* les droites passant par un couple de rendements successifs et par l'origine) passent par un des quatre points cardinaux ou leurs déclinaisons (32 directions au total, voir graphique A.1). Ainsi, on parlera de droites « bien placées » quand elles se superposeront avec un des axes cardinaux (*Cf.* graphique A.1 en annexe pour une illustration). Dans le cas d'une rose des vents « parfaite », nous n'aurons que des « droites bien placées », passant par l'origine et rayonnant dans toutes les directions cardinales. A l'opposé, plus le nombre de droites

« mal placées » sera important et moins le phénomène de rose des vents sera présent. Le phénomène est alors bien caractérisé par le nombre de droites « mal placées », d'où la formulation suivante du premier indicateur $CRII_1$. Il est défini à partir du rapport entre le nombre de droites caractérisant parfaitement une rose des vents et le nombre de droites passant par chacun des couples de rendements et l'origine.

Formellement, nous avons :

$$CRII_1 = 4 / \{ \kappa^{-1} [\#(D_0) - 4] + 4 \} \quad (1.1)$$

où D_0 désigne l'ensemble des droites reliant les points et passant par l'origine, $\#(D_0)$ désigne le nombre d'éléments de D_0 , et κ (avec $0 < \kappa < 1$) une constante représentant la pénalité pour les droites mal placées ; le nombre $[\#(D_0) - 4]$ représente ainsi le nombre de droites mal placées⁶.

Théoriquement, si le phénomène de rose des vents existe dans la série considérée, l'indice $CRII_1$ doit tendre vers sa valeur maximale⁷.

1.2.2.2 Un deuxième indice d'intensité du phénomène de rose des vents ($CRII_2$)

A la suite des travaux de Koppl et Nardone (2001), Vorlow (2004) et Antoniou et Vorlow (2005), nous avons construit un deuxième indice appelé *Compass Rose Intensity Index 2* ($CRII_2$). L'idée sous-jacente est que le phénomène de rose des vents est caractérisé par la présence de plusieurs droites reliant les différents points du *Scatter plot* des rendements successifs, et passant toutes par l'origine. A partir de ce constat, Koppl et Nardone (2001) proposent de transformer les couples de rendements (r_t, r_{t-1}) en leurs coordonnées

⁶La notion de droites « mal placée » ici fait référence aux droites ne passant pas par un point cardinal ou une de ses déclinaisons. Ainsi, la droite ayant une pente de -1 et celle avec une pente de 1/2 n'auront pas la même influence sur la valeur de l'intensité.

⁷Pour faciliter les comparaisons ultérieures (Cf. infra), nous avons de plus imposé de manière *ad hoc* une normalisation supplémentaire des différents indicateurs $CRII$, en utilisant une fonction appliquée à l'indicateur brut (Cf. Abramowitz et Stegun, 1964), définie par : $f(x) = 1 - \exp(-\lambda x^3)$, avec λ fixé entre 0 et 1, et x la valeur de l'indicateur à normaliser. Ainsi, l'indicateur $CRII_1$ (comme les autres indicateurs) aura une valeur finie comprise dans l'intervalle $[0, 100]$.

polaires (R_t, ω_t) . Soient :

$$R_t = \sqrt{r_t^2 + r_{t-1}^2} \quad (1.2)$$

et

$$\omega_t = \theta_t / \pi \quad (1.3)$$

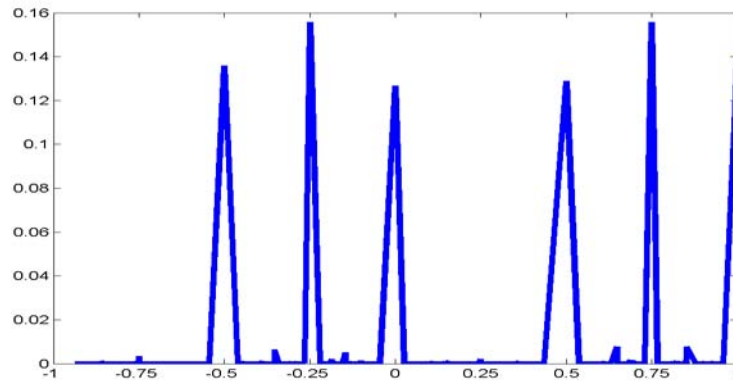
avec :

$$\theta_t = \begin{cases} \arctan(r_t/r_{t-1}) & \text{si } (r_{t-1} \geq 0) \\ \arctan(r_t/r_{t-1}) + \pi & \text{si } (r_{t-1} < 0) \text{ et } (r_t \geq 0) \\ \arctan(r_t/r_{t-1}) - \pi & \text{si } (r_{t-1} < 0) \text{ et } (r_t < 0) \end{cases}$$

et $-\pi/2 \leq \arctan(x) \leq \pi/2$ pour tout $x \in \mathbb{R}$.

En comptabilisant les couples de rendements dont les secondes coordonnées polaires θ_t (normalisées par π) sont similaires (dans l'intervalle $[-\delta\theta_t, +\delta\theta_t]$ où δ est une constante qui dépend de la précision des rendements calculés), nous obtenons, d'après la terminologie de Koppl et Nardone (2001), un histogramme de *thetas* (*Empirical Theta Histogram*, voir graphique 1.6), qui permet de mettre en lumière les distorsions liées à l'effet de discrétisation amplifiant le phénomène sous-jacent.

Cet histogramme permet, de manière simple, de rendre compte de la dispersion des valeurs de la seconde coordonnée polaire normalisée et on peut comparer l'histogramme empirique à celui d'une loi uniforme pour mettre en évidence un effet de discrétisation. On peut aussi isoler le phénomène de rose des vents en comparant la distribution des coordonnées polaires de la série sous-jacente avec celle d'une *pseudo-série* obtenue en tirant au hasard des rendements dans la série originale (avec remise) et en les associant de manière aléatoire selon une loi uniforme (procédure de *bootstrap*). Dans ce deuxième

FIG. 1.6. Histogramme des valeurs normalisées du θ 

Source : simulation de 30 000 points à partir d'un modèle $AR(2) - GARCH(1,1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1,1)$ sont respectivement de $\{5, 2.10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2.10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondante est ensuite déterminée par discrétisation des prix en suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR et un pas d'échelon de cotation de 1 EUR. L'axe horizontal représente les fréquences relatives; l'axe horizontal représente les différentes valeurs des angles normalisés.

cas de figure, on pourra isoler l'effet de corrélation de l'effet de discrétisation. Basé sur ces considérations, le deuxième indice d'intensité de rose des vents proposé s'écrit :

$$CRII_2 = \sum_{w=-1}^1 \left\{ \frac{[\hat{f}(w) - u(w)]^2}{u(w)} \right\} + \sum_{w=-1}^1 \left\{ \frac{[\hat{f}(w) - \hat{g}(w)]^2}{\hat{g}(w)} \right\} \quad (1.4)$$

où $\hat{f}(\cdot)$ est la fonction de densité empirique des secondes coordonnées polaires normalisées notées w , $u(\cdot)$ la loi uniforme et $\hat{g}(\cdot)$ la fonction de densité empirique associée à la série des couples de rentabilités ré-arrangées aléatoirement.

L'indice prendra de grandes valeurs s'il existe un effet de discrétisation des prix car la première partie de l'indice sera élevée; l'indice sera aussi élevé quand la structure de dépendance temporelle de la série diffère nettement de l'hypothèse d'indépendance tem-

porelle (Cf. Brock *et alii*, 1996), ce qui contribuera à augmenter la seconde partie de l'indice⁸.

1.2.2.3 Un troisième indice d'intensité du phénomène de rose des vents ($CRII_3$)

L'idée du troisième indice est la suivante : le phénomène se manifeste par la présence des couplets dans des parties spécifiques du diagramme de phases. Si l'on représente ce diagramme sous la forme d'un tableau en discrétisant les axes (voir graphique A.2 en annexe), alors certaines parties de ce tableau devraient être plus remplies que d'autres. Ainsi, en considérant tous les couples de rendements successifs, il est possible de construire un tableau de contingence. On peut obtenir une mesure similaire du phénomène directement à partir des rendements eux-mêmes et non plus à partir de leurs coordonnées polaires. Là encore, on peut aussi essayer de distinguer l'effet de discrétisation de l'effet de dépendance temporelle en comparant le tableau de contingence empirique avec ceux issus d'une loi uniforme et de la série « *bootstrappée* ». En utilisant la norme de Frobenius⁹ calculée sur les différences des matrices de contingence, l'indice $CRII_3$ s'écrit :

$$CRII_3 = \|z\|^2 = \sum_{i=1}^k \sum_{j=1}^k (\hat{p}_{ij} - u_{ij})^2 + \sum_{i=1}^k \sum_{j=1}^k (\hat{p}_{ij} - \tilde{p}_{ij})^2 \quad (1.5)$$

avec : \hat{p}_{ij} la fréquence empirique des rendements se trouvant dans la cellule (i, j) du tableau de contingence de la série brute ; u_{ij} la fréquence correspondant à une loi uniforme ; \tilde{p}_{ij} la fréquence correspondant à la série indépendante (« *bootstrappée* » à partir des valeurs de la série originale) et $k = Ent[4 \times 0.75(n-1)^2]^{\frac{1}{5}}$ le nombre *optimal* de partition

⁸L'indicateur $CRII_2$ s'inspire directement d'un test traditionnel du Khi-deux (Cf. Greenword et Nikulin, 1996), qui compare des valeurs observées à des valeurs simulées. Nous notons ici que le problème est bien posé numériquement, puisque la valeur de la variable u , issue d'une loi uniforme, sera toujours différente de zéro. De même, la deuxième partie de l'indicateur (effet de corrélation) est elle aussi bien définie dans la mesure où la variable g est issue d'un tirage aléatoire (procédure de bootstrap) toujours différente de la valeur (strictement) nulle. On peut voir sur le graphique 1.7 (plus bas dans le texte, simulation des prix sans structure de corrélation) que, quel que soit le pas d'échelon de cotation fixé, le phénomène n'est pas visible et que l'indicateur $CRII_2$ rend bien compte de cette réalité avec une valeur faible. Par contre, lorsqu'une structure de corrélation est imposée lors des simulations (voir graphique 1.8 plus bas dans le texte), le phénomène s'intensifie avec la taille de l'échelon de cotation.

⁹Voir Ledoit et Wolf (2003 et 2004) pour l'utilisation de cette norme dans le cadre d'une optimisation robuste de portefeuille.

des rendements, avec n le nombre d'éléments du vecteur de rentabilité.

Ainsi, comme dans le cas de l'indice $CRII_2$, cet indicateur fournira de grandes valeurs s'il existe un effet de discrétisation des prix car la première partie de l'indice sera élevée ; il sera aussi élevé quand la structure de dépendance temporelle de la série diffèrera nettement de l'hypothèse d'indépendance temporelle, ce qui contribuera à augmenter la seconde partie de l'indice.

1.2.2.4 Complémentarités et limites des mesures d'intensité proposées

Les trois précédentes mesures permettent chacune isolément de quantifier le phénomène dans la majorité des cas, mais, dans des situations ambiguës, elles seront néanmoins complémentaires. En effet, si l'on considère un cas certes extrême où tous les points sont alignés sur l'axe des ordonnées, alors l'indicateur $CRII_1$ aura une valeur très élevée signalant faussement la présence d'un phénomène de rose des vents ; par contre $CRII_2$ et $CRII_3$ vont permettre de rejeter cette présence du phénomène car chacun aura une valeur faible. De plus, la troisième mesure complémentaire $CRII_3$, basée sur le nombre de cases vides d'un tableau de contingence, signalera, autrement que visuellement, des situations où les droites rayonnantes ne seront pas assez étirées pour rendre compte d'un « vrai » phénomène de rose des vents.

Les mesures proposées dans cette section ont pour vocation principale d'aider à caractériser de manière plus formelle le phénomène de rose des vents *via* le calcul d'indicateurs quantitatifs simples. A cet égard, nous montrons par simulations et sur données réelles qu'ils caractérisent bien les situations de présence indéniable du phénomène. Les résultats obtenus, dans la suite de cet article, montrent que le phénomène a de très grandes chances d'être pleinement caractérisé si les trois indicateurs sont relativement élevés (supérieurs à 60), alors qu'au contraire, ceux-ci sont faibles (inférieurs à 20) lorsque l'existence du phénomène ne peut être invoquée. Entre ces deux limites extrêmes, le

phénomène, par définition continu, existe mais est plus ou moins visible¹⁰.

1.3 Applications et caractérisations empiriques du phénomène

Après avoir défini l'intensité du phénomène de rose des vents à partir des trois indices précédents, nous cherchons dans la section suivante à mettre en évidence les conditions d'existence de ce phénomène, et par cette voie détournée, certains des éléments nécessaires à son apparition. Nous commencerons d'abord par étudier le phénomène à partir d'une série artificielle constituée de tous les rendements possibles évalués sur un intervalle de prix donné; puis nous envisagerons le phénomène avec des données simulées à partir de processus de rendements bien connus. Nous testerons enfin, à partir de données de marché, l'existence et l'ampleur du phénomène sur le marché français.

1.3.1 Conditions d'existence sur données simulées

Nous allons, à partir de simulations des couples de rendements possibles (sans corrélation dans un premier temps, ensuite avec corrélation connue et enfin à partir des processus connus de la littérature sur la modélisation des données financières), analyser les conditions nécessaires et suffisantes d'existence d'un phénomène de rose des vents. Ce qui nous permettra par la suite de valider les hypothèses sous-jacentes à la construction des indices d'intensité de rose des vents.

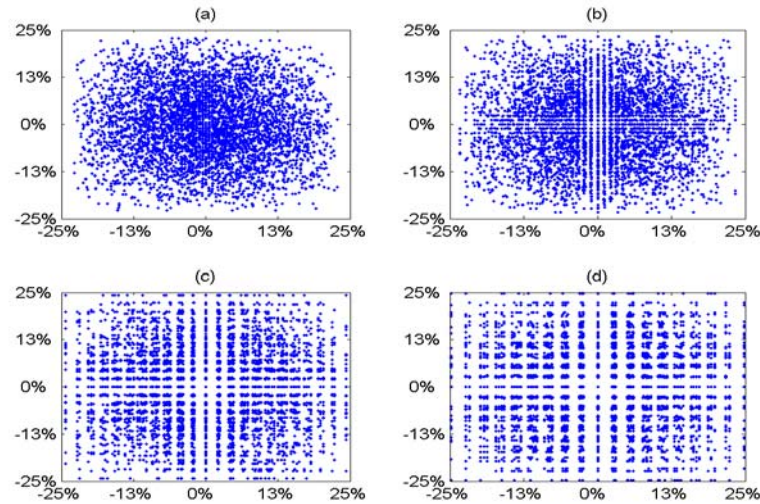
1.3.1.1 Corrélation temporelle des rendements nulle

¹⁰Alors que le phénomène est très présent en temps de transactions, celui-ci tend à disparaître à partir d'une agrégation à une fréquence de trente minutes (voir graphique A.3 en annexe, pour lequel les valeurs des trois indicateurs sont respectivement égales à 75, 48 et 27 en temps de transactions, contre 0, 1 et 3 pour des cotations fréquentées à 30 minutes).

Considérons un intervalle arbitraire de prix, par exemple $[80, 120]$, et discrétisons celui-ci en définissant un pas de cotation (par exemple 1 EUR). Calculons ensuite par combinaisons (extensives) l'ensemble des rentabilités possibles à partir de l'intervalle de prix initial. Finalement, associons de manière aléatoire avec une loi uniforme deux rendements et représentons la série de couples de rendements obtenus. Nous avons alors, dans le cas d'une série indépendante, l'ensemble des couples de rendements qu'il sera possible de voir apparaître sur un graphique xy (*Scatter Plot*). Le graphique suivant représente ce cas de figure pour quatre différentes tailles d'échelon de cotation.

L'analyse visuelle de ces graphiques et les différents indices *CRII* ne permettent pas, même s'il existe un effet de discrétisation des prix, de mettre en évidence un phénomène de rose des vents, quand le pas de l'échelon de cotation - relatif ou absolu - augmente. Néanmoins, une structure apparaît quand la taille de l'échelon augmente et se traduit par des bandes de couples de rendements impossibles, plus larges au centre qu'aux extrémités du graphique. Autrement dit, la discrétisation des cours a un effet plus net sur les faibles variations de cours et on retrouve ici l'analyse de Crack et Ledoit (1996). Mais cette structure n'est pas conforme aux caractéristiques *stricto sensu* d'une rose des vents, quelle que soit la volatilité autorisée du sous-jacent (c'est-à-dire l'étendue de l'intervalle arbitraire considéré). L'échelon de variation, sa taille relative ou absolue, et la volatilité du rendement du sous-jacent ne sont pas des déterminants individuels suffisants pour expliquer la présence d'une rose des vents. Il faut donc une dépendance entre les rendements. Le paramètre crucial concernant l'apparition d'une rose des vents pourrait donc être la corrélation du premier ordre des rentabilités.

FIG. 1.7. Phénomène de rose des vents pour différentes tailles d'échelon de cotation (sans corrélation)



Source : simulation de 30 000 points à partir de tous les cas possibles de prix dans l'intervalle [80, 120]. La série de rendements arrondis correspondante est ensuite déterminée par discrétisation des prix en suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR. L'axe horizontal représente les rendements du titre à une période donnée; l'axe vertical représente les rendements à la période précédente. Les graphiques (a), (b), (c) et (d) représentent le *scatter-plot* des rendements pour différentes tailles de l'échelon de cotation (voir le tableau ci-dessous où d représente la taille de l'échelon et où les indices $CRII$ sont les mesures d'intensité du phénomène).

	d	$CRII_1$	$CRII_2$	$CRII_3$
(a)	0,5	0,08	0,02	0,11
(b)	1,0	0,08	0,09	0,10
(c)	2,0	0,11	0,27	0,17
(d)	2,5	0,13	0,45	0,28

1.3.1.2 Structures de corrélation compatibles avec une *pseudo*-rose de vents

Pour qu'une rose des vents se forme, il faut que les données exhibent une structure de corrélation particulière. Une des treize manières de représenter une corrélation entre des variables présentées par Rodgers et Nicewander (1988) repose précisément sur les angles formés par chacun des points du graphique xy (appelé *Person Space*). Une structure « pure » de rose des vents sera matérialisée par un ensemble très réduit de coefficients de

corrélation dans l'espace de *Person*, soit $\{-1, 0, 1\}$; une structure moins claire comportera un ensemble plus large de coefficients de corrélation ($\{-0,75; -0,25; 0; 0,25; 0,75\}$ par exemple), alors que l'absence de tout phénomène de rose des vents se traduira par une distribution uniforme des coefficients de corrélation (cas limite). De même, plusieurs processus générateurs de prix avec des dépendances sérielles existent dans la littérature financière. Nous présentons les résultats des simulations pour ces différents cas de figure.

Structure ad hoc

Pour étudier les déterminants d'une rose des vents, commençons par imposer une structure particulière de corrélation. Pour construire les graphiques suivants, nous avons simulé à partir d'une série complète de rendements possibles sur une plage de variation donnée (ici $[80, 120]$), un ensemble de couples de rendements (R_{t+1}, R_t) ordonnancés de la manière suivante :

$$R_{t+1} = \tilde{\rho} R_t + \varepsilon_t \quad (1.6)$$

avec $\tilde{\rho}$ un coefficient de corrélation tiré aléatoirement de manière équiprobable dans l'ensemble $\{-1; -0,75; -0,25; 0; 0,25; 0,75; 1\}$ et ε_t un bruit blanc distribué suivant une loi normale¹¹.

La série des prix est ensuite obtenue par l'application de la fonction exponentielle sur la série précédente avec la valeur initiale de 80. De façon plus précise, si P_t désigne le prix de l'actif à la période t , nous avons :

$$P_t = P_{t-1} \exp(r_t) \quad (1.7)$$

avec $P_0 = 80$.

¹¹Le rendement initial R_0 est ici fixé de façon arbitraire à 2%. Cette valeur correspond approximativement à la moyenne des rendements maximum des actifs considérés dans l'échantillon (Cf. supra). Voir à cet effet le tableau 1.2 pour les statistiques descriptives des rendements des titres considérés.

Pour tenir compte des conditions de discrétisation, nous avons arrondi les prix obtenus à partir de l'équation (1.7). La méthode est celle utilisée par Fang (2002). Supposons que le prix observé à la date t est P_t^* , alors :

$$P_t^* = Ent(P_t/d + 0.5)d \quad (1.8)$$

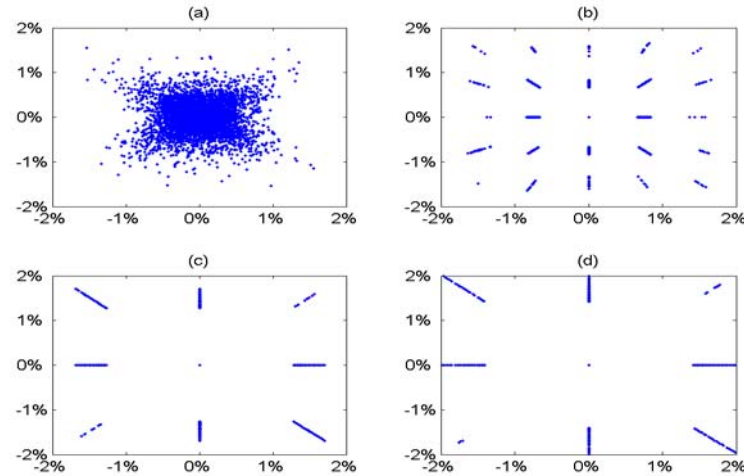
où la fonction $Ent(x)$ renvoie le plus grand entier inférieur ou égal à x et le paramètre d représente la taille de l'échelon de cotation.

Dans la suite de cet article, nous faisons varier l'échelon de cotation et analysons le phénomène de rose des vents. On voit clairement sur les graphiques suivants que la rose des vents apparaît, pour un nombre de tirage donné, à partir d'un certain seuil de pas de cotation (relatif et absolu). Nous retrouvons là encore les résultats de Crack et Ledoit (1996).

La littérature ne met pas néanmoins en lumière l'effet de la taille de l'échantillon sur l'existence du phénomène. Si celui-ci est clairement présent dès que l'on a suffisamment de points dans l'échantillon, il a aussi tendance, *ceteris paribus*, à disparaître quand la taille de l'échantillon augmente. Nous expliquons ce résultat par la présence de plus en plus importante dans l'échantillon de grandes valeurs de couples de rendements. Nous avons vu en effet précédemment que la taille de l'échelon de cotation avait beaucoup d'influence sur la discrétisation des rendements quand ceux-ci étaient relativement petits. Autrement dit, c'est la réalisation des valeurs extrêmes et leurs présences dans l'échantillon quand celui-ci devient grand, qui tend à défaire la structure de rose des vents. On retrouve là la même influence sur le phénomène que la volatilité. Pour un même échelon de cotation, plus la série est volatile et plus les couples de rendements importants vont se trouver dans la partie de l'échantillon échappant à l'effet de discrétisation (*magnet effect*, Szpiro, 1998). Plus l'échantillon de cours considéré sera grand et plus le nombre de valeurs extrêmes

dans les couples de rendements considérés sera important et moins l'effet de discrétisation se fera sentir.

FIG. 1.8. Phénomène de rose des vents pour différentes tailles d'échelon de cotation (avec une structure de corrélation imposée)



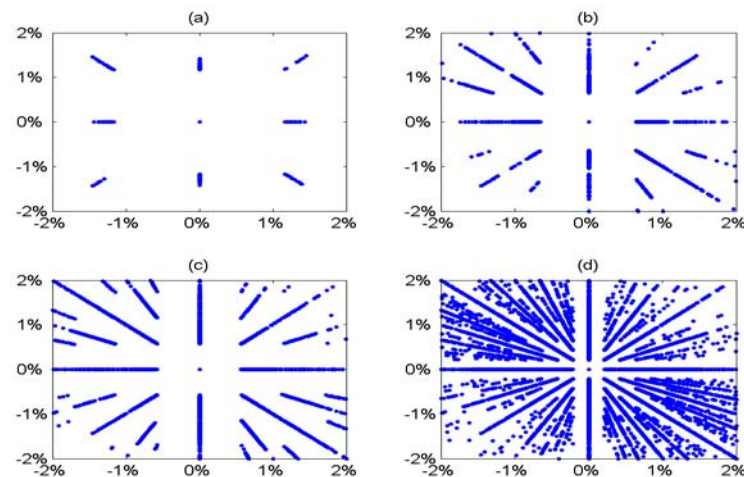
Source : simulation de 30 000 points à partir de l'équation (1.6) qui représente la relation de corrélation temporelle entre les séries. L'espace de corrélation est fixé à $[0, 0,25, 0,75, -0,25, -0,75]$. La série de rendements arrondis correspondante est ensuite déterminée par discrétisation des prix en suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR. L'axe horizontal représente les rendements du titre à une période donnée; l'axe vertical représente les rendements à la période précédente. Les graphiques (a), (b), (c) et (d) représentent le *scatter-plot* des rendements pour différentes tailles de l'échelon de cotation (voir le tableau ci-dessous où d représente la taille de l'échelon et où les indices $CRII$ sont les mesures d'intensité du phénomène).

	d	$CRII_1$	$CRII_2$	$CRII_3$
(a)	0,01	2,77	0,02	0,60
(b)	0,50	13,33	32,39	21,23
(c)	1,00	100,00	56,32	50,79
(d)	1,50	100,00	67,48	62,65

Pour illustrer ce phénomène nous représentons dans le graphique suivant les indices $CRII$ obtenus pour l'échantillon des couples de rendements se trouvant dans l'ellipse des 65% les plus probables et l'on constate que la taille relative de l'échantillon n'a plus d'importance.

Cela illustre bien l'effet des valeurs extrêmes des rendements qui tend à désagréger la structure de rose des vents.

FIG. 1.9. Diagramme de phases des rentabilités pour différentes tailles de l'échantillon



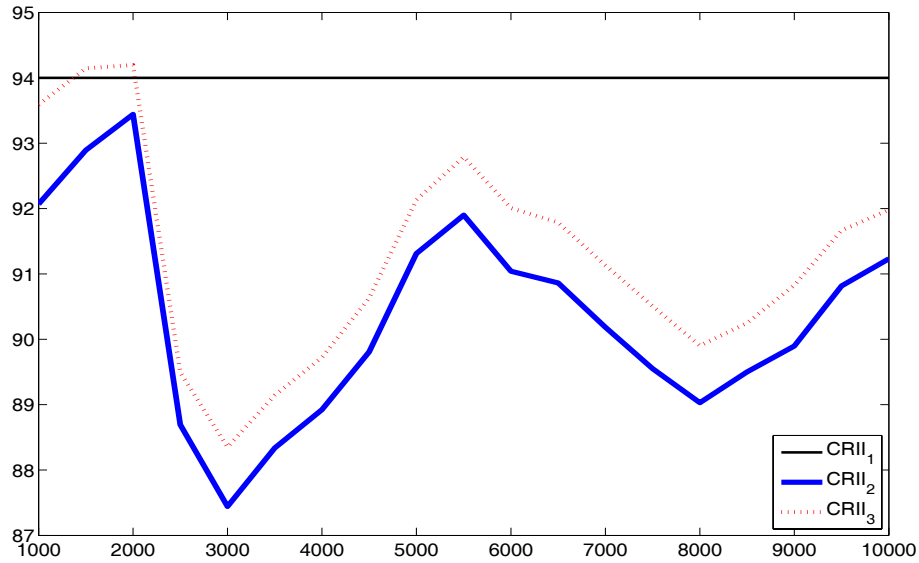
Source : simulation à partir d'un modèle $AR(2) - GARCH(1, 1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1, 1)$ sont respectivement de $\{5, 2.10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2.10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondant est ensuite déterminée par discrétisation des prix suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR et un pas d'échelon de cotation de 1 EUR. L'axe horizontal représente les rendements du titre à une période donnée ; l'axe vertical représente les rendements à la période précédente. Les axes sont exprimés en %. Les graphiques (a), (b), (c) et (d) représentent le *scatter-plot* des rendements pour différentes tailles de l'échantillon (voir le tableau ci-dessous où n représente la taille de l'échantillon, σ la volatilité et où les indices $CRII$ sont les mesures d'intensité du phénomène).

	n	σ	$CRII_1$	$CRII_2$	$CRII_3$
(a)	500	0,02082	100,00	90,02	80,60
(b)	1 000	0,02012	63,33	62,39	61,23
(c)	5 000	0,01992	33,05	26,32	30,79
(d)	10 000	0,01878	10,01	7,48	5,65

Structure ex ante

Adoptons maintenant la démarche similaire à celle de la précédente section - dans laquelle nous imposons une structure de corrélation finale aux données - en imposant cette fois

FIG. 1.10. Indices d'intensité de rose des vents pour différentes tailles de l'échantillon



Source : simulation à partir d'un modèle $AR(2) - GARCH(1, 1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1, 1)$ sont respectivement de $\{5, 2 \cdot 10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2 \cdot 10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondant est ensuite déterminée par discrétisation des prix suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR et un pas d'échelon de cotation de 1 EUR. L'axe horizontal représente la taille de l'échantillon ; l'axe vertical représente le niveau des mesures d'intensité du phénomène de rose des vents.

une structure particulière au processus générateur des données et en les faisant passer au filtre de la discrétisation des données lié à la présence d'un pas de cotation. Plusieurs modèles sur les processus sous-jacents ont été proposés dans la littérature et nous avons choisi ceux parmi les plus populaires pour l'analyse des données financières. Les modèles $GARCH$, qui sont parmi les plus utilisés pour la modélisation (Bollerslev, 1986 et Bollerslev *et alii*, 1992), permettent la description de l'évolution endogène de la volatilité. Ainsi, nous simulons un $AR(2) - GARCH(1, 1)$ à la Bollerslev (1986), un modèle $GARCH$ avec le processus de bruit qui suit une loi de *Student*, un modèle $GARCH$ avec le processus de bruit qui est distribué selon un *Skew-Normal* (Valle, 2004) et un modèle $GARCH$ avec le processus de bruit suivant une loi *Skew-Student* (Keith et Shen, 2004) en utilisant

certaines valeurs estimées des paramètres déjà reportées dans la littérature (Cf. Amilon et Byström, 2000).

- Modèle *GARCH*

Un processus *GARCH* (*Generalized AutoRegressive Conditional Heteroskedasticity*) défini par Bollerslev (1986) permet de rendre compte de la persistance de la volatilité. De façon analytique, un modèle *GARCH*(p, q) se présente comme suit :

$$r_t = \varepsilon_t \sqrt{\sigma_t^2} \quad (1.9)$$

avec :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

où r_t représente le rendement, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\forall (i, j) \leq (q, p) \in \mathbb{N}^2$, σ_t^2 est la variance conditionnelle de ε_t , avec ε_t un bruit blanc distribué suivant une loi normale.

- Modèle *Skew-Normal* et *Skew-t*

Les modèles *Skew-Normal* (notés *SN*, Cf. Valle, 2004) et *Skew-t* (notés *ST*, Cf. Keith et Shen, 2004) font partis de la classe des distributions de probabilité qui généralisent la distribution gaussienne avec des paramètres régulant le degré d'asymétrie et les queues de distribution. Ces modèles vont nous permettre d'étudier les familles d'actifs non-symétriques avec des queues de distribution épaisses.

Dans ce cas, un processus x issu d'une loi *Skew-Normal*, avec un paramètre d'échelle α (représentant la mesure de l'asymétrie), avec $\alpha \in \mathbb{R}$, a une densité de probabilité se présentant comme suit :

$$g(x|\alpha) = 2\phi(x)\Phi(\alpha x) \quad (1.10)$$

avec $x \in \mathbb{R}$ et où $\phi(\cdot)$ et $\Phi(\cdot)$ sont respectivement la fonction de densité et la fonction de

répartition de la loi normale, soient :

$$\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$$

$$\text{et} \quad \Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x \exp(-u^2/2) du$$

De façon générale, si $y = \xi + \sigma x$, avec $\xi \in \mathbb{R}$ (paramètre de niveau représentant la moyenne de y) et $\sigma > 0$ (volatilité de y) ; si $y \sim SN(\xi, \sigma, \alpha)$, alors sa fonction de densité s'écrit :

$$g(y^* | \xi, \sigma, \alpha) = (\pi\sigma)^{-1} \exp(-y^{*2}/2) \int_{-\infty}^{\alpha y^*} \exp(-u^2/2) du \quad (1.11)$$

avec $y^* = (y - \xi)/\sigma$.

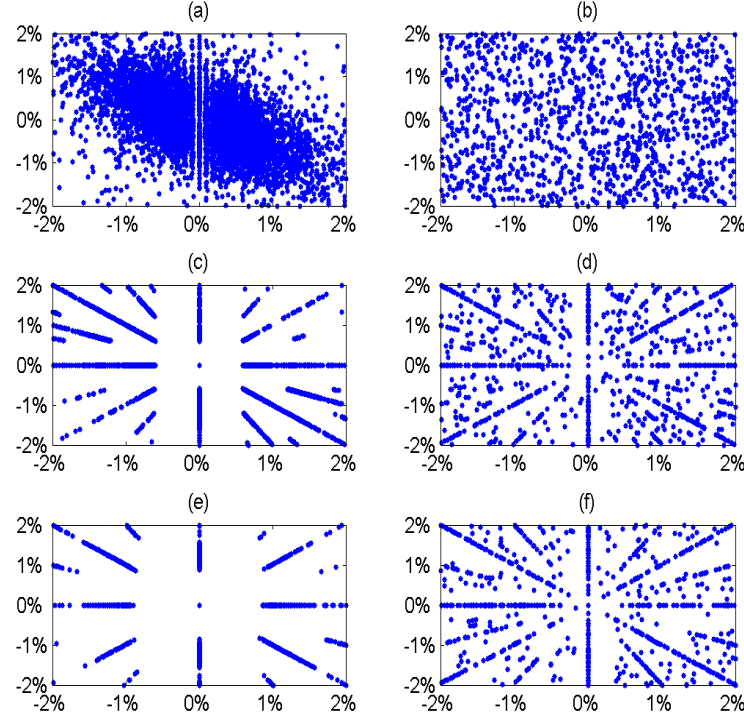
De même, si z est une variable aléatoire issue d'une loi *Skew-t* de paramètre α , avec $\alpha \in \mathbb{R}$, alors sa densité de probabilité vérifie l'équation (1.10) dans laquelle les fonctions $t(\cdot)$ et $T(\cdot)$ respectivement les fonctions de densité et de répartition de la loi de *Student* se substituent à leurs analogues de la loi normale. En d'autres termes, si $z \sim ST(\xi, \sigma, \nu, \alpha)$ où ν est le degré de liberté, alors sa fonction de densité s'écrit :

$$f(z^* | \xi, \sigma, \nu, \alpha) = 2t(z^*, \nu)T(z^*Q | \nu + 1) \quad (1.12)$$

avec : $Q = \alpha[(1+\nu)/(z^{*2}+\nu)]^{1/2}$, $z^* = (z - \xi)/\sigma$, et $t(\cdot)$ et $T(\cdot)$ respectivement la fonction de densité et la fonction de répartition de la loi de *Student*.

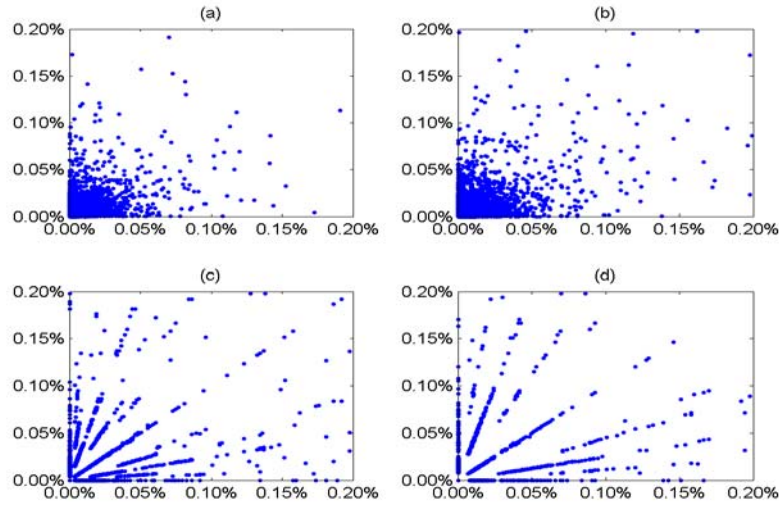
Nous représentons dans les graphiques 1.11 à 1.13 les effets isolés des déterminants d'un phénomène de rose des vents (quand les autres paramètres se situent dans des plages moyennes de variation).

FIG. 1.11. Diagramme de phases des rentabilités pour différentes tailles de l'échelon et de volatilité de la série - modèle *AR-GARCH*-



Source : simulation de 30 000 points à partir d'un modèle $AR(2) - GARCH(1,1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1,1)$ sont respectivement de $\{5, 2 \cdot 10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2 \cdot 10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondant est ensuite déterminée par discrétisation des prix suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR. L'axe horizontal représente les rendements du titre à une période donnée; l'axe vertical représente les rendements à la période précédente. Les axes sont exprimés en 10^{-3} . Les graphiques (a), (b), (c), (d), (e) et (f) représentent le *scatter-plot* des rendements pour différentes tailles de l'échelon de cotation et de volatilité (voir le tableau ci-dessous où d représente la taille de l'échelon, σ la volatilité et où les indices $CRII$ sont les mesures d'intensité du phénomène).

	d	σ	$CRII_1$	$CRII_2$	$CRII_3$
(a)	0,01	0,001176	1,77	0,02	0,60
(b)	0,01	0,010796	0,33	0,39	1,23
(c)	0,10	0,001449	80,30	76,32	78,79
(d)	0,10	0,013237	60,73	57,98	52,69
(e)	0,50	0,001788	91,00	86,32	80,79
(f)	0,50	0,013226	67,02	60,48	62,65

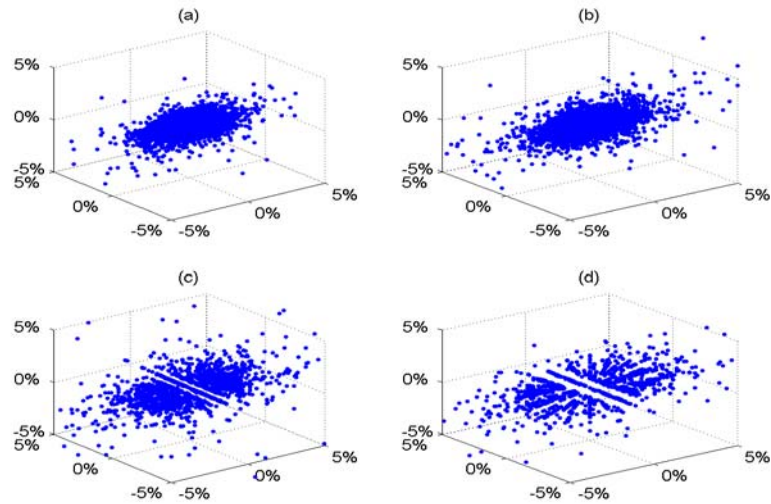
FIG. 1.12. Eventail selon l'échelon de cotation - modèle *AR-GARCH* -

Source : simulation de 30 000 points à partir d'un modèle $AR(2) - GARCH(1,1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1,1)$ sont respectivement de $\{5, 2 \cdot 10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2 \cdot 10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondant est ensuite déterminée par discrétisation des prix suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR. L'axe horizontal représente les rendements du titre au carré à une période t donnée; l'axe vertical représente les rendements au carré à la période précédente. Les axes sont exprimés en 10^{-3} . Les graphiques (a) (b) (c) et (d) représentent le *scatter-plot* des rendements pour différentes tailles de l'échelon de cotation (voir le tableau ci-dessous où d représente la taille de l'échelon, σ la volatilité et où les indices *CRII* sont les mesures d'intensité du phénomène).

	d	σ	$CRII_1$	$CRII_2$	$CRII_3$
(a)	0,01	0,00010	3,77	1,52	2,10
(b)	0,10	0,00110	11,30	17,39	19,21
(c)	0,50	0,00027	56,10	56,32	50,79
(d)	1,00	0,00033	70,00	67,48	62,65

Le premier paramètre considéré pour les simulations est la taille de l'échantillon et le second le pas relatif de cotation. On remarque sur les graphiques précédents que les cinq hypothèses considérées pour le processus générateur sont compatibles avec le phénomène de rose des vents pour des pas de cotation réalistes. Ainsi, on constate que le phénomène est peu visible pour une taille de l'échelon de cotation faible et devient nettement visible au fur et à mesure que la taille de l'échelon de cotation augmente (*Cf.* Graphiques 1.11, 1.13 et 1.12). Ceci est tout à fait conforme aux conclusions de Crack et Ledoit (1996). Que ce soit sur le graphique 1.14 ou sur les graphiques 1.15 et 1.16, on remarque que le phénomène semble disparaître quand la taille de l'échantillon devient élevée. En effet, pour une taille de 1 000, le phénomène de rose des vents est bien visible et le nombre de droites différentes partant de l'origine est très faible. Par contre, ce nombre de droites est plus élevé lorsque la taille de l'échantillon passe à 5 000, 10 000 ou 30 000 comme on peut le voir sur les graphiques 1.14 à 1.16. L'explication peut venir du fait que lorsque la taille de l'échantillon augmente, le nombre de points « aberrants » (de taille anormale) augmente aussi.

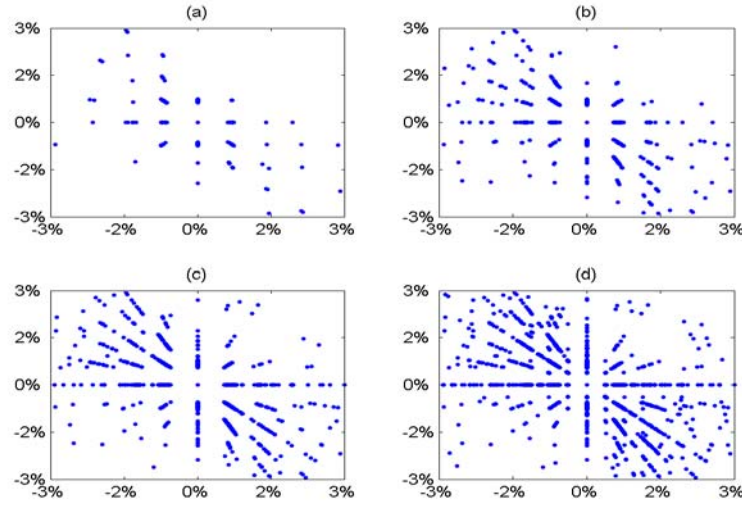
Ainsi, quelle que soit la procédure de simulation et la loi postulée, un phénomène de rose des vents apparaît dans les séries de rendements lorsque les prix sont discrétisés, ce qui confirme donc les principales conclusions des études antérieures sur le sujet.

FIG. 1.13. Phénomène d'explosion d'étoiles - modèle $AR-GARCH$ -

Source : simulation de 30 000 points à partir d'un modèle $AR(2) - GARCH(1,1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1,1)$ sont respectivement de $\{5, 2 \cdot 10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2 \cdot 10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondant est ensuite déterminée par discrétisation des prix suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR. L'axe horizontal représente les rendements du titre à une période t donnée; l'axe horizontal représente les rendements du titre à une période $t - 1$ donnée; l'axe vertical représente les rendements à la période $t - 2$. Les graphiques (a) (b) (c) et (d) représentent le *scatter-plot* des rendements pour différentes tailles de l'échelon de cotation (voir le tableau ci-dessous où d représente la taille de l'échelon, σ la volatilité et où les $CRII$ sont les mesures d'intensité du phénomène).

	d	σ	$CRII_1$	$CRII_2$	$CRII_3$
(a)	0,01	0,00784	12,77	10,02	10,60
(b)	0,10	0,00784	23,33	19,39	29,23
(c)	0,50	0,00816	40,00	46,32	40,79
(d)	1,00	0,01095	70,00	77,48	72,65

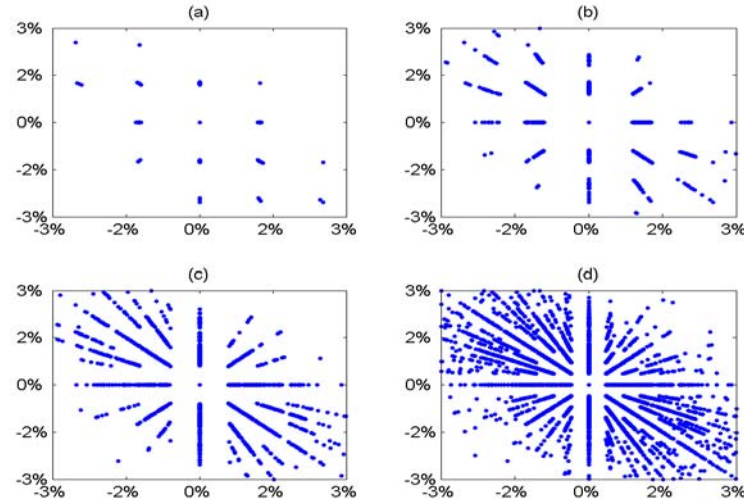
FIG. 1.14. Diagramme de phases des rentabilités pour différentes tailles de l'échantillon
- modèle *AR-Student-GARCH* -



Source : simulation à partir d'un modèle $AR(2) - T(4) - GARCH(1,1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1,1)$ sont respectivement de $\{5, 2.10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2.10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondant est ensuite déterminée par discrétisation des prix suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR et un pas d'échelon de cotation de 1 EUR. L'axe horizontal représente les rendements du titre à une période donnée ; l'axe vertical représente les rendements à la période précédente. Les graphiques (a) (b) (c) et (d) représentent le *scatter-plot* des rendements pour différentes tailles de l'échantillon (voir le tableau ci-dessous où n représente la taille de l'échantillon, σ la volatilité et où les indices $CRII$ sont les mesures d'intensité du phénomène).

	n	σ	$CRII_1$	$CRII_2$	$CRII_3$
(a)	500	0,02216	62,77	59,02	50,60
(b)	1 000	0,02038	43,63	42,13	41,23
(c)	5 000	0,02676	30,07	26,32	25,17
(d)	10 000	0,02536	10,90	7,98	11,76

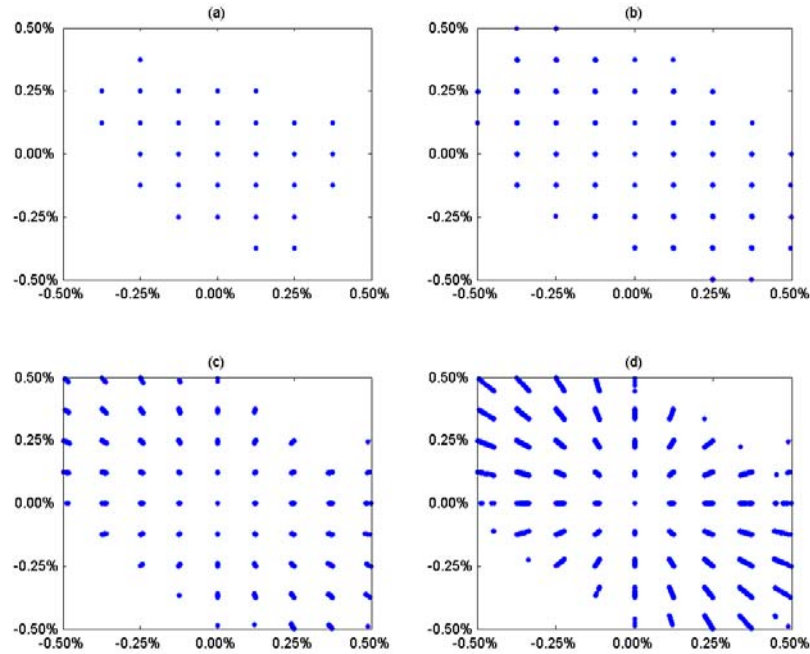
FIG. 1.15. Diagramme de phases des rentabilités pour différentes tailles de l'échantillon - modèle *AR-Skew-Normal-GARCH* -



Source : simulation à partir d'un modèle $AR(2) - SN(0,006, 1, 0, 0151) - GARCH(1, 1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1, 1)$ sont respectivement de $\{5, 2.10^{-4}, 0, 12, 0, 053\}$ et $\{2, 2.10^{-5}, 0, 11, 0, 83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondant est ensuite déterminée par discrétisation des prix suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR et un pas d'échelon de cotation de 1 EUR. L'axe horizontal représente les rendements du titre à une période donnée; l'axe vertical représente les rendements à la période précédente. Les axes sont exprimés en %. Les graphiques (a) (b) (c) et (d) représentent le *scatter-plot* des rendements pour différentes tailles de l'échantillon (voir le tableau ci-dessous où n représente la taille de l'échantillon et où les indices $CRII$ sont les mesures d'intensité du phénomène).

	n	$CRII_1$	$CRII_2$	$CRII_3$
(a)	500	72,57	60,72	50,69
(b)	1 000	53,65	52,90	47,80
(c)	5 000	30,10	26,52	30,99
(d)	10 000	20,00	23,11	18,65

FIG. 1.16. Diagramme de phases des rentabilités pour différentes tailles de l'échantillon - modèle *AR-Skew-Student-GARCH* -



Source : simulation à partir d'un modèle $AR(2) - ST(0,002, 0,33, 3, 0,0151) - GARCH(1,1)$; les paramètres du modèle $AR(2)$ et ceux du $GARCH(1,1)$ sont respectivement de $\{5, 2.10^{-4}, 0, 12, 0,053\}$ et $\{2, 2.10^{-5}, 0, 11, 0,83\}$ (paramètres obtenus par Amilon et Byström, 2000, pour les actions « Atlas Copco A fria » sur le marché financier de Stockholm). La série de rendements arrondis correspondant est ensuite déterminée par discrétisation des prix suivant les procédures (1.7) et (1.8) avec un prix initial fixé à 80 EUR et un pas d'échelon de cotation de 0,5 EUR. L'axe horizontal représente les rendements du titre à une période donnée ; l'axe vertical représente les rendements à la période précédente. Les axes sont exprimés en %. Les graphiques (a) (b) (c) et (d) représentent le *scatter-plot* des rendements pour différentes tailles de l'échantillon (voir le tableau ci-dessous où n représente la taille de l'échantillon et où les indices $CRII$ sont les mesures d'intensité du phénomène).

	n	$CRII_1$	$CRII_2$	$CRII_3$
(a)	500	2,77	0,02	0,60
(b)	1 000	13,33	32,39	21,23
(c)	5 000	10,00	56,32	50,79
(d)	10 000	10,00	57,48	52,65

1.3.2 Conditions d'existence sur données de marché

Après avoir construit une mesure quantitative du phénomène et caractérisé ces manifestations à l'aide de données simulées dans le cadre de modèles réalistes, nous présentons une étude du phénomène sur quelques titres représentatifs du marché français.

1.3.2.1 La base de données

La base de données étudiée contient les séries de prix de haute fréquence (données de transactions) de quelques titres (Alcatel, Alstom, Axa, BNP, France Telecom, Michelin, Peugeot et Sanofi) de la place boursière de Paris sur le premier semestre 2003 pour Alcatel et le premier semestre 2004 pour les autres actifs (*Cf.* tableau 1.1 suivant). Nous avons en moyenne pour chaque titre un total d'environ un million d'observations. L'écart maximum (*Range*) est faible pour tous les titres. Quelques caractéristiques principales des prix sont présentées dans le tableau 1.1 suivant.

TAB. 1.1. Caractéristiques des prix des titres considérés

Titres	Min.	Max.	Période	Nombre de points
Alcatel	4,1	4,9	03/01/2003 - 13/07/2003	261 117
Alstom	29,2	33,6	02/01/2004 - 13/08/2004	1 655 129
Axa	127	140,9	02/01/2004 - 14/08/2004	2 910 942
BNP	85	94,2	02/01/2004 - 13/08/2004	2 834 980
France Telecom	116	136	02/01/2004 - 01/07/2004	4 354 910
Michelin	36,7	41,9	02/01/2004 - 16/08/2004	929 420
Peugeot	208	245	02/01/2004 - 13/08/2004	1 283 125
Sanofi	36,1	42	02/01/2004 - 13/08/2004	1 812 028

Source : Euronext et calculs des auteurs. Données de transactions.

1.3.2.2 Résultats empiriques sur le marché français des actions

Les cours étant mesurés à très haute fréquence, c'est-à-dire transaction-par-transaction, la taille des séries sur une même période peut être très différente d'un actif à un autre. En effet, certains titres très liquides peuvent faire l'objet de plusieurs transactions pendant une minute, alors que d'autres, moins liquides, auront peu de transactions. Cette base est appropriée pour prendre en compte l'ensemble d'informations disponibles sur les séries.

Elle nous permet dans la suite d'analyser et de faire ressortir l'effet de la discrétisation des prix sur les plages de rentabilités.

Le tableau 1.2 fournit un descriptif des séries de rendements des différents titres étudiés pour mettre en évidence, à travers nos indices, le phénomène de rose des vents sur le marché français des actions. Le tableau 1.3 fournit les auto-corrélations d'ordre 1 des rendements ainsi que des rendements carrés, cubes et puissance 4 (*Cf.* Avouyi-Dovi *et alii*, 2002) ainsi que quelques statistiques représentatives des durations (*Cf.* Engle et Russell, 1997). On peut remarquer sur le tableau 1.3 que la fréquence moyenne des échanges est inférieure à 1 minute pour la plupart des titres étudiés.

TAB. 1.2. Tableau descriptif des rendements des titres de l'échantillon

Titres	Min.	Max.	Moyenne (10^{-6})	Ecart-type (%)	Skewness	Kurtosis
Alcatel	-0,015	0,025	4,4	0,2	0,24	8,74
Alstom	-0,063	0,055	-8,5	0,29	0,18	50,55
Axa	-0,029	0,014	-6	0,14	-0,84	25,42
BNP	-0,013	0,017	-4,99	0,1	0,17	20,2
France Telecom	-0,067	0,009	-1,02	0,14	-9,92	488,07
Michelin	-0,015	0,03	-4,81	0,19	0,54	17,72
Peugeot	-0,029	0,018	1,09	0,19	-0,03	16,02
Sanofi	-0,026	0,032	-7,57	0,19	0,3	27,17

Source : Euronext et calculs des auteurs. Données de transactions. Les calculs sont effectués sur le premier semestre 2003 pour Alcatel et sur le premier semestre 2004 pour les autres titres (*Cf.* Tableau 1.1).

Nous reproduisons aussi les auto-corrélogrammes de tous les titres en annexe du document - graphique A.3 - qui mettent en évidence une relation négative entre deux rendements successifs, qui disparaît quand l'échantillonnage est produit à 30 minutes.

A partir de ces séries de rendements, nous pouvons constater (voir le graphique 1.17) que le phénomène de rose des vents est bien présent sur la plupart des titres du marché français des actions, comme une simple analyse visuelle permet de le constater. En effet, sur l'ensemble des titres étudiés, nous pouvons, à partir d'un diagramme de phases (gra-

TAB. 1.3. Auto-corrélations d'ordre 1 des rendements (et transformés) et statistiques descriptives sur les durations

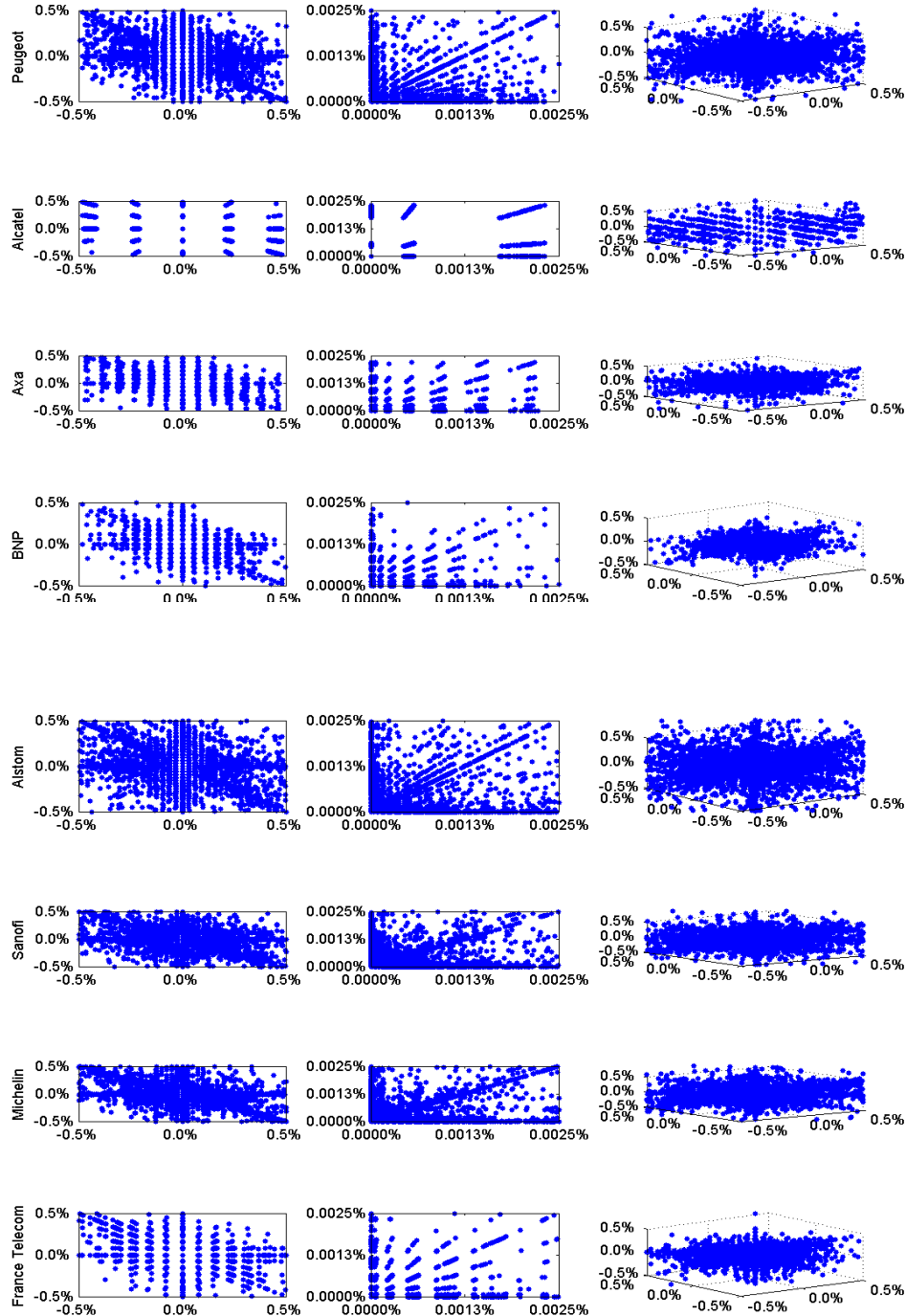
Titres	ρ	$\rho^{Vol.}$	$\rho^{skew.}$	$\rho^{kurt.}$	μ des durations (en secondes)	σ des durations (en secondes)
Alcatel	-0,423*	0,404*	-0,023*	0,097*	20	178
Alstom	-0,398*	0,421*	-0,415*	0,407*	59	318
Axa	-0,357*	0,183*	-0,039*	0,011*	34	235
BNP	-0,372*	0,268*	-0,069*	0,017*	34	238
France Telecom	-0,319*	0,021*	-0,002*	0,001*	22	189
Michelin	-0,397*	0,343*	-0,110*	0,031*	105	434
Peugeot	-0,354*	0,205*	-0,029*	0,003*	76	363
Sanofi	-0,357*	0,339*	-0,149*	0,062*	54	301

Source : Euronext et calculs des auteurs. Données de transactions. Les calculs sont effectués sur le premier semestre 2003 pour Alcatel et sur le premier semestre 2004 pour les autres titres (Cf. Tableau 1.1). Le signe * indique une significativité au seuil de 1% (Cf. Campbell *et alii*, 1997).

phique 1.17, première colonne), remarquer l'émergence de plusieurs droites passant par l'origine et fuyant vers les points cardinaux (rose des vents). On voit notamment que les points sont concentrés autour de quatre droites principales : l'axe des abscisses, l'axe des ordonnées, la première bissectrice et la deuxième bissectrice. Cette remarque est confirmée par l'analyse des graphiques des rentabilités au carré (graphique 1.17, deuxième colonne) où l'on peut voir une figure typique d'éventail pour toutes les séries et par l'analyse des graphiques des rentabilités dans un espace de dimension trois (graphique 1.17, troisième colonne) où le phénomène d'explosion d'étoiles est nettement observé sur tous les titres étudiés.

L'intensité de cette structure est néanmoins nettement différente d'un titre à l'autre selon que les points sont bien ou mal placés sur les quatre droites principales caractérisant le phénomène de rose des vents. Les résultats de l'analyse visuelle précédente sont confirmés par les trois différentes mesures proposées pour quantifier cette structure. En effet, comme le montre le tableau 1.4 suivant, les intensités ($CRII_1$ et $CRII_2$) sont d'autant plus fortes que le phénomène de rose des vents est nettement visible.

FIG. 1.17. Rose des vents, éventail et explosion d'étoiles des différents titres cotés sur le marché français des actions



Source : Euronext et calculs des auteurs. Pour les graphiques de la première colonne (*Scatter plot*), l'axe horizontal représente les rendements du titre à une période donnée; l'axe vertical représente les rendements à la période précédente. Pour les graphiques de la deuxième colonne (éventail), l'axe horizontal représente les rendements au carré du titre à une période donnée; l'axe vertical représente les rendements au carré à la période précédente. Enfin, pour les graphiques de la troisième colonne (explosion d'étoiles), le premier axe horizontal représente les rendements du titre à une période t donnée; le second axe horizontal représente les rendements du titre à une période $t-1$ donnée; l'axe vertical enfin représente les rendements à la période $t-2$.

Elle est plus nette pour Alcatel, France Telecom et Axa ; par contre elle l'est moins pour Alstom, Sanofi et Michelin. Quant à la mesure $CRII_3$, même si les valeurs sont moins élevées que celles des deux indices précédents, les trois résultats sont similaires. Il faut noter que cette intensité n'exprime pas la qualité de la représentation visuelle mais plutôt le degré de dépendance sérielle à travers la distribution des points dans un tableau de contingence *optimal*. A cet égard, la structure est bien plus uniformément répartie pour les titres France Telecom, Alstom et Axa par rapport à BNP, Peugeot et Michelin.

TAB. 1.4. Intensité du phénomène de rose des vents pour les titres considérés

Titres	Inverse de la volatilité	CRII1(%)	CRII2(%)	CRII3(%)	Visibilité de la rose des vents
Alcatel	499,22	60,61	74,48	15,9	Elévée
Alstom	342,22	30,53	5,6	22,03	Faible
Axa	723,49	69,77	40,14	18,69	Elévée
BNP	994,29	60	40,31	11,29	Moyenne
France Telecom	735,29	75,47	48,37	27,45	Elévée
Michelin	514,24	25,47	8,32	14,22	Faible
Peugeot	538,07	60,61	13,88	11,77	Moyenne
Sanofi	515,86	38,09	7,39	18,06	Faible

Source : Euronext et calculs des auteurs. Données de transactions. Les calculs sont effectués sur le premier semestre 2003 pour Alcatel et sur le premier semestre 2004 pour les autres titres (*Cf.* Tableau 1.1).

Ainsi, à partir d'une valeur des indicateurs supérieure à 60%, la rose des vents est bien caractérisée et visible sur un graphique, alors que pour des valeurs inférieures à 20% le phénomène n'est pas détectable sur un graphique et il ne peut ainsi être invoqué. Entre ces deux valeurs limites, le phénomène existe mais n'est pas bien caractérisé.

On retrouve donc à travers les trois mesures complémentaires les résultats de Crack et Ledoit (1996). En effet, la discrétisation des prix des titres sur un marché est à l'origine de la structure complexe de la distribution des rentabilités dans un diagramme de phases. Toutefois, les effets de cette structure peuvent être limités à partir d'un seuil d'échantillonnage de 25 à 30 minutes en fonction de l'actif.

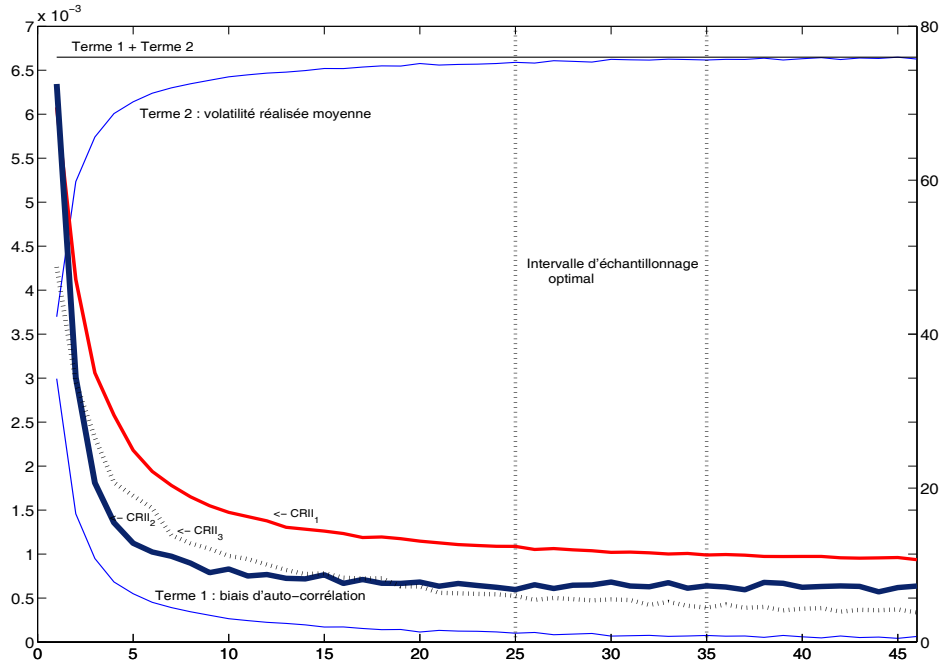
1.4 Conclusion

Partant des remarques et considérations générales de l'article séminal de Crack et Ledoit (1996), nous avons analysé le phénomène de rose des vents dans son ensemble. Nous avons d'abord montré que celui-ci apparaissait uniquement lorsque les prix des actifs étaient discrets et que la discrétisation des prix avait un effet (relatif et absolu) sur les variations significatives des prix. Ensuite, nous avons constaté que le phénomène de rose des vents n'était visible qu'à partir d'un certain seuil du ratio taille de l'échelon de cotation sur volatilité de la série. Ce résultat confirme les conclusions de Crack et Ledoit (1996), Lee *et alii* (2005), McKenzie et Frino (2003), ou de Wang *et alii* (2000) qui montrent l'existence de seuils d'apparition du phénomène.

Notre apport principal concerne la construction d'indicateurs d'intensité du phénomène qui permettent - sauf cas pathologiques - de bien caractériser la structure. Les indices d'intensité sont en effet nuls lorsque les prix des actifs sont parfaitement continus et élevés lorsque ceux-ci sont discrétisés avec une forte taille d'échelon de cotation effective. De plus, ces indicateurs montrent la relation positive qui existe entre le phénomène de rose des vents et le ratio taille de l'échelon de cotation sur volatilité de la série des rendements. Nous avons illustré de manière simple le rôle de la taille de l'échantillon et mis en avant l'hypothèse que ce sont les grandes variations de prix - peu présentes dans les petits échantillons - qui ont tendance à perturber la netteté du phénomène. Nous avons aussi montré que quelques processus utilisés dans la littérature financière sont tout à fait compatibles avec le phénomène de rose des vents. Autrement dit, ces modèles simples rendent compte suffisamment finement des caractéristiques financières - et en particulier de leur dépendance temporelle - pour ne pas être mis en défaut par ce phénomène avéré de microstructure. Après avoir mis en évidence le phénomène sur le marché français en considérant quelques uns des titres les plus liquides de la place, nous avons enfin déterminé les seuils d'agrégation temporelle pour lesquels le phénomène tend à disparaître (valeurs des indicateurs inférieures à 20).

Cette partie est à rapprocher des travaux de Gottlieb et Kalay (1985) ou de Amilon et Byström (2000) qui s'interrogent sur les effets de la présence de ce type de phénomène sur les propriétés des estimateurs des paramètres caractéristiques de modèles de séries temporelles, et de Oomen (2002) et Griffin et Oomen (2008) qui montrent qu'il existe un seuil d'agrégation *optimal* pour mesurer la volatilité réalisée (*Realized Volatility*) en minimisant le biais de mesure lié à la dépendance sérielle des rendements de haute fréquence. Ce dernier résultat est illustré par le graphique 1.18 qui montre la relation décroissante entre le pas d'échantillonnage et la qualité des estimateurs, dans lequel nous voyons que le phénomène disparaît, tout comme le biais d'auto-corrélation, pour des fréquences plus basses que 25-30 minutes.

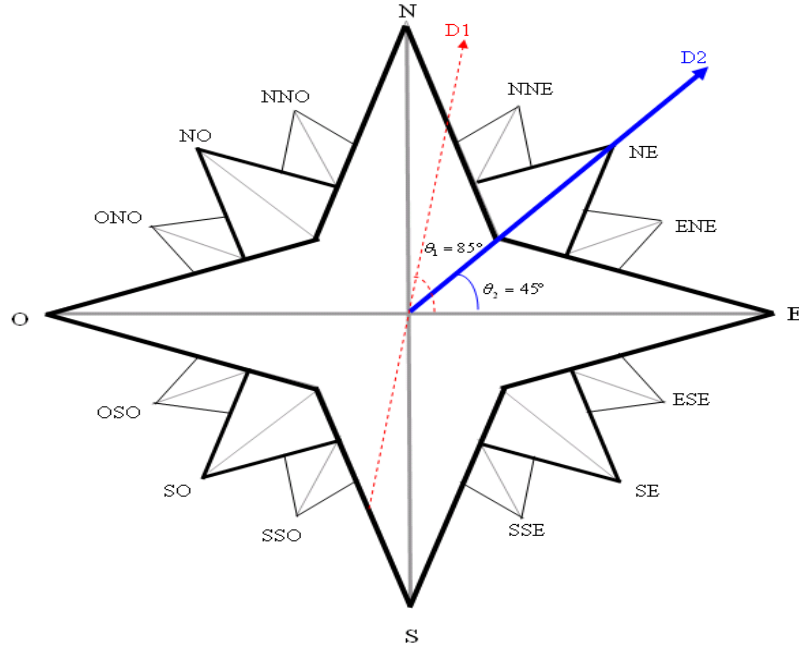
FIG. 1.18. Moyennes des volatilités réalisées, des biais d'auto-corrélation et des mesures de la rose des vents en fonction de la fréquence d'échantillonnage



Source : Euronext. Données de transactions allant du 03/01/2003 au 13/07/2003 pour Alcatel et du 02/01/2004 au 16/08/2004 pour les autres titres. Le graphique présente, échelle de gauche, la moyenne normalisée des volatilités réalisées journalières (qui sont les moyennes des rendements carré intra journaliers, *i.e.* $RV = T^{-1} \sum_{t=1}^T \sum_{i=1}^{N_h} R_{h,t,i}^2$) des huit titres considérés dans l'étude, ainsi que le biais moyen sur le facteur d'auto-corrélation d'ordre 1 (calculé par : $Biais = 2T^{-1} \sum_{t=1}^T \sum_{i=1}^{N_h-1} \sum_{j=1}^{N_h} R_{h,t,i} R_{h,t,j}^2$), pour des fréquences d'échantillonnage entre 1 et 46 minutes (*Cf.* Oomen, 2002); et les indicateurs de rose des vents sur l'axe de droite. Pour cette représentation graphique et à l'instar de Oomen (2002), la variance globale sur la période (Terme 1 + Terme 2) et la variance réalisée sont ici calculées, alors que le biais d'auto-corrélation est déduit de ces deux précédentes valeurs.

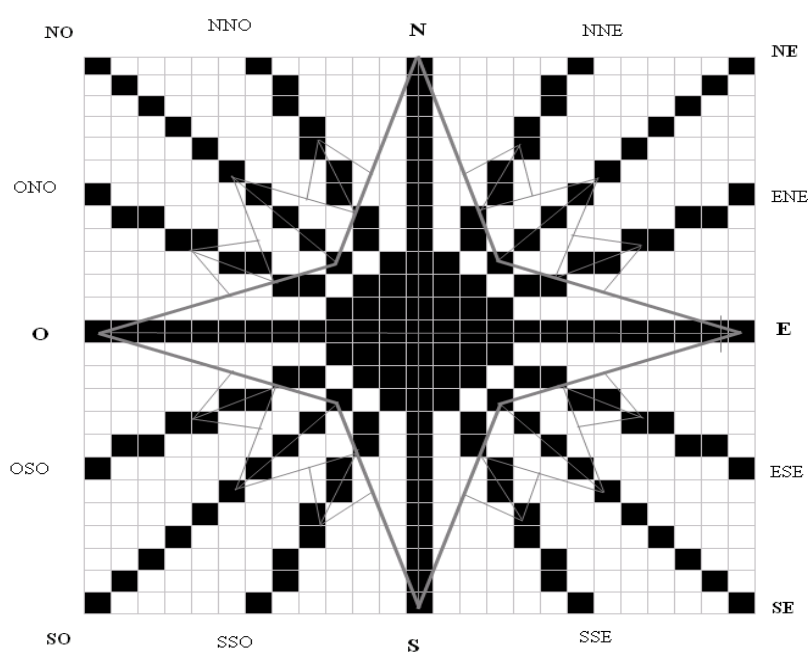
Annexes : graphiques complémentaires

FIG. A.1. Illustration de la caractérisation de droites « bien ou mal placées » pour l'indicateur $CRII_1$



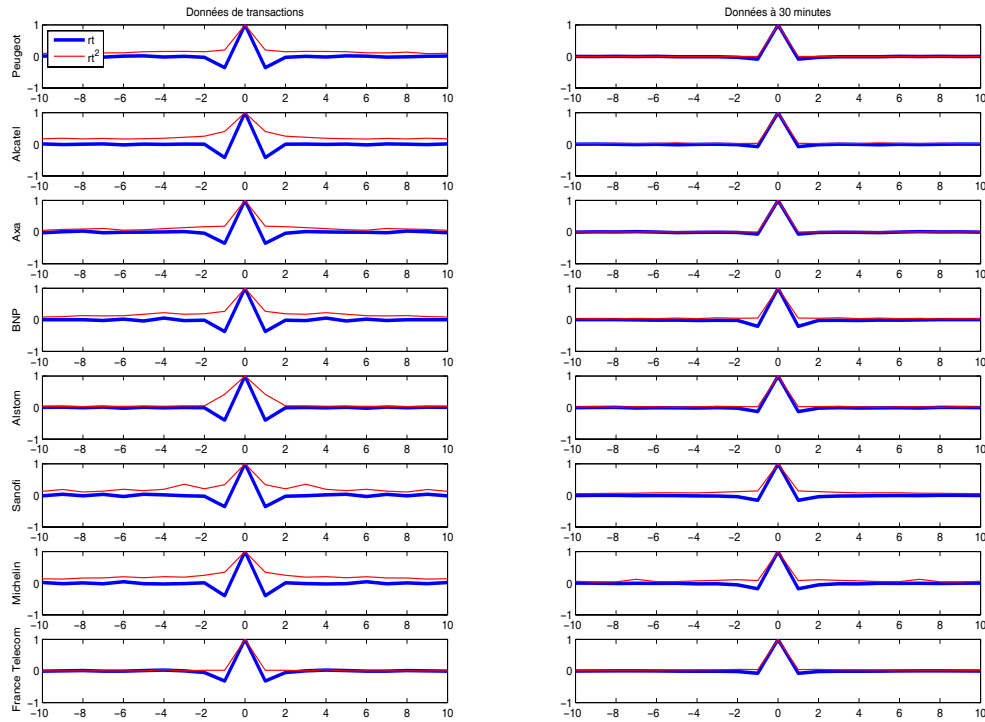
Note : les sigles N, S, O, E, NE, NO, SE, SO, NNE, NNO, SSE, SSO, ONO, OSO, ENE, ESE représentent respectivement, le Nord, le Sud, l'Ouest, l'Est, le Nord-Est, le Nord-Ouest, le Sud-Est, le Sud-Ouest, le Nord-Nord-Est, le Nord-Nord-Ouest, le Sud-Sud-Est, le Sud-Sud-Ouest, l'Ouest-Nord-Ouest, l'Ouest-Sud-Ouest, l'Est-Nord-Est et l'Est-Sud-Est. La droite D2 dessinée ci-dessus (en trait plein gras) est dite « bien placée » du fait de son alignement sur l'axe Sud-Ouest Nord-Est. Par contre, la droite D1 (en pointillés) est dite « mal placée » car elle ne passe pas par un des axes principaux d'une rose des vents.

FIG. A.2. Illustration de la caractérisation par pixels de la rose des vents pour l'indicateur $CRII_3$



Note : les sigles N, S, O, E, NE, NO, SE, SO, NNE, NNO, SSE, SSO, ONO, OSO, ENE, ESE représentent respectivement, le Nord, le Sud, l'Ouest, l'Est, le Nord-Est, le Nord-Ouest, le Sud-Est, le Sud-Ouest, le Nord-Nord-Est, le Nord-Nord-Ouest, le Sud-Sud-Est, le Sud-Sud-Ouest, l'Ouest-Nord-Ouest, l'Ouest-Sud-Ouest, l'Est-Nord-Est et l'Est-Sud-Est. Les cases noires du damier ci-dessus correspondent aux cellules du tableau de contingence dans lesquelles des couples de rendements existent.

FIG. A.3. Auto-corrélations d'ordre 1 à 10 de quelques titres du marché français en données de transactions et fréquences à 30 minutes



Source : Euronext. Données de transactions allant du 02/01/2004 au 16/08/2004. Les graphiques représentent les auto-corrélations d'ordre -10 à 10 des séries de rendements et du carré des rendements (*proxy* de la volatilité) de quelques titres du marché français des actions. Voir le tableau 1.4 pour les valeurs des indicateurs de rose des vents. Les graphiques de gauche, sur données de transactions ont des auto-corrélations au premier ordre des rendements et des rendements carrés statistiquement significatives au seuil de 5%, tandis que ceux de droite, sur données cadencées à 30 minutes, rendent compte d'auto-corrélations qui ne sont pas significatives au même seuil.

Chapter 2

High-frequency Market Risk

Scenarii by Independent

Component Analysis: Method and Application

2.1 Introduction

Value-at-Risk (VaR) represents the loss over a pre-set time horizon with a given probability. This risk measure is used for risk control and management, including both credit and operational risks (Duffee and Pan, 1997, 2001 and Tsai, 2004). The objective of VaR computation is to provide an unbiased assessment of the portfolio risk over time. To achieve this objective, the framework must be flexible and general enough to accommodate changes of the portfolio due to the changes in the associated risk factors. The VaR concept is popular since the 1996 amendments to the Basel Capital Accord incorporating market risk into the regulatory capital framework. Supervisory agencies allow banks to use their internal VaR models to calculate their regulatory capital requirements as an alternative to the conventional methodology. Another reason that VaR was embraced by

banks and regulators is that it offers a measure of a firm's overall exposure to market risk, which is useful for understanding how much a given instrument or portfolio might lose over a period of time. Despite several critics, many observations recognize that the systematic use of the VaR has an overall positive effect on the handling of risk within financial institutions, matching the increasing degree of complexity of investment instruments over the last few years.

Financial literature also focuses on the use of intra-day data for better measurements of risk (see Dacorogna *et al.*, 2001 and Rydberg and Shephard, 2003), more accurate asset pricing models and robust asset allocation (see Bollerslev and Zhang, 2003). Finally, traditional models in finance are based on the existence of global factors driving the market evolution (see Ross, 1976 and Gençay *et al.*, 2003).

Indeed, there are probably many factors that cause the movements of asset returns whose evolutions are rather complex and unpredictable. It is not unusual to assume that different assets are driven by common factors with special characteristics. Loretan (1997) makes use of this idea by extracting the main statistical factors applying a Principal Component Analysis (PCA) to the asset returns in a financial context. Once these factors backed-out from market data, and due to the orthogonality of the Principal Components, estimation and simulation are straightforward, yielding easy generation of virtual portfolio trajectories. Loss of information, however, may be entailed by the linear filter of the PCA when high-frequency financial data is considered. Non-linearities and non-gaussianities are indeed specific features of this type of data (see McNeil and Frey, 2000; Voit, 2003; Härdle *et al.*, 2003). In this case, an Independent Component Analysis (ICA, see Comon, 1994 and Hyvärinen *et al.*, 2001) seems to be a more appropriate technique as mentioned by Back and Weigend (1997). The ICA, a well-known method for finding latent structure in data, is a statistical model that expresses a set of multidimensional observations as a combination of unknown latent variables. These underlying latent variables are called sources or Independent Components, statistically independent

to each other *per* construction. We use this technique when generating *scenarii* reduces the effective dimensionality of the *scenario* specification problem, without leaving aside important characteristics of the data.

The chapter is organized in five sections; first section presents illustrations of VaR computations related to high-frequency stock returns. Section 2 presents the main characteristics of the ICA whilst section 3 focuses on VaR computations. Section 4 contains the empirical results and section 5 concludes.

2.2 Independent Component Analysis

The method of Independent Component Analysis (ICA) briefly summarized in this section, is well-known in Signal Processing and applied in various fields: biomedicine, speech and telecommunications signals to mention a few. Signals can be seen as mixtures of different physical sources and external noise sources. The goal here is to find the independent sources. We next present the formal definition of the method and then justify its use in a financial time-series analysis framework.

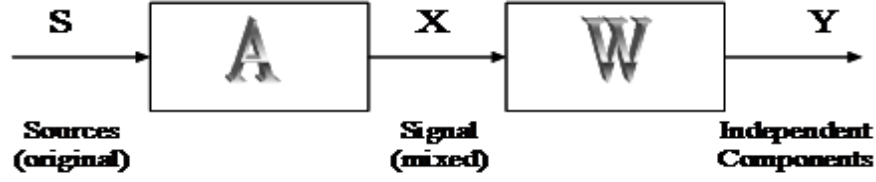
2.2.1 Definition

The ICA (see Comon, 1994 and Hyvärinen *et al.*, 2001) is a statistical method that expresses a set of multidimensional observations as a combination of unknown latent variables. These underlying latent variables are called sources or Independent Components and are assumed to be statistically independent of each other. The schematic representation of the ICA model is in figure 2.1.

Figure 2.1 shows the most basic form of the ICA. The original sources are mixed through matrix \mathbf{A} to form the observed signal \mathbf{x} . The demixing matrix \mathbf{W} transforms the observed signal into the independent components \mathbf{Y} . Therefore, we can write:

$$\mathbf{x} = f(\beta, \mathbf{s}) \tag{2.1}$$

Figure 2.1. Schematic Representation of ICA



where \mathbf{x} is an observed vector with size $(m \times 1)$ and $f(\cdot)$ is a general unknown function, with parameters β , that operates on statistically independent latent variables listed in the vector \mathbf{s} with size $(n \times 1)$. A special case of relation (2.1) is obtained when the function is linear, and thus:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (2.2)$$

where \mathbf{A} is an unknown $(m \times n)$ mixing matrix. In *formulae* (2.1) and (2.2), \mathbf{x} and \mathbf{s} are random vectors. When a sample of observations $\mathbf{X} = (x_1, \dots, x_N)$ becomes available, we have $\mathbf{X} = \mathbf{A}\mathbf{S}$, where the matrix \mathbf{X} has observations \mathbf{x} as its columns and similarly, the matrix \mathbf{S} has latent variable vectors \mathbf{s} as its columns. The mixing matrix \mathbf{A} is supposed to be constant for all observations.

The mixing process is hidden, so we can only observe the mixed signals. The objective is to recover the original source signals from the observations through a demixing process. The following equations (2.3) and (2.4) mathematically describe the demixing processes:

$$\mathbf{Y} = \mathbf{W}\mathbf{X} \quad (2.3)$$

and so:

$$\mathbf{Y} = \mathbf{W}\mathbf{A}\mathbf{S} \quad (2.4)$$

Matrix \mathbf{S} contains the original source signals driving the observations whereas the separated signals are stored in matrix \mathbf{Y} . The $(n \times m)$ matrix \mathbf{W} is the demixing matrix. If the separated signals are the same as the original sources ($\mathbf{Y} = \mathbf{S}$), the mixing matrix is the inverse of demixing matrix ($\mathbf{A} = \mathbf{W}^{-1}$).

In general, since both \mathbf{S} and \mathbf{A} are unknown, any scalar multiple in one of the sources can always be cancelled by dividing the corresponding column of \mathbf{A} by this scalar (because fixed magnitudes of ICs assuming unit variance; *i.e.*: $E(\mathbf{S}_{\cdot j}^2) = 1, \forall j \leq n$). Therefore, the ICA cannot determine the variances of the IC's, $\forall i \leq \mathbf{N}$:

$$\mathbf{x}_i = \sum_{j=1}^n \mathbf{a}_{ij} \mathbf{s}_j = \sum_{j=1}^n \frac{\mathbf{a}_{ij}}{\mathbf{c}_j} (\mathbf{c}_j \mathbf{s}_j), \quad (2.5)$$

with $E(\mathbf{s}_j^2) = E[(\mathbf{c}_j \mathbf{s}_j)^2] = 1$, $\mathbf{x}_i = \mathbf{X}_i$ and $\mathbf{s}_j = \mathbf{S}_{\cdot j}$

For the same reason, ICA cannot determine the order of the ICs (terms can be freely changed):

$$\mathbf{X} = \mathbf{AS} = \mathbf{AP}^{-1}\mathbf{PS} \quad (2.6)$$

where \mathbf{P} is a permutation matrix and \mathbf{PS} are the original IC's but in a different order.

Various ICA algorithms have been proposed in the literature to find the demixing \mathbf{W} . To find such a matrix \mathbf{W} , the following assumptions are made (Back and Weigend, 1997):

- The sources \mathbf{s} are statistically independent. While this might appear to be strong hypothesis, it is not an unreasonable assumption when one considers, for example, sources of very different origins ranging from foreign politics to microeconomic variables that might impact a stock price.
- At most one source has a Gaussian distribution. In the case of financial data, pure and independently normally distributed signals are so rare that only allowing for

one of them is not a serious restriction.

- The signals are stationary. Stationarity is an heroic but classical assumption for the main model.

2.2.2 Estimation of the ICA model

A standard approach for ICA algorithms is the following two-step procedure (Cardoso, 1999 and Cardoso and Souloumiac, 1993): Decorrelation (or Whitening) and Minimization. In the first step, we seek to diagonalize the covariance matrix of the input signals. The second step minimizes a measure of the higher order statistics that ensuring the non-gaussian output signals are as statistically independent as possible. It can be shown that this can be carried out by an unitary rotation matrix (see Cardoso and Souloumiac 1993). This second step provides the higher order independence.

The task is to find both the latent variables or sources \mathbf{s}_j and the mixing process; in the linear case, the latter task consists of finding the mixing matrix \mathbf{A} . A popular approach is to find a separating matrix \mathbf{W} so that variables \mathbf{y}_j in relation 2.3 are the estimation of \mathbf{s}_j up to scaling and permutation. Hence \mathbf{W} is an estimate of the pseudo-inverse of \mathbf{A} up to scaling and permutation of the rows of \mathbf{W} . Often the latent variables \mathbf{s}_j are estimated one-by-one, by finding a column vector \mathbf{w}_j (this will be stored as a row of \mathbf{W} ; *i.e.*: $\mathbf{w}_j = \mathbf{W}_{j.}$) such that $\mathbf{y}_j = \mathbf{w}_j \mathbf{X}$ is an estimation of \mathbf{s}_j .

There are several approaches for estimating the Independent Components and the Mixing Matrix, resulting in different algorithms. In all approaches, an objective (or a contrast) function $G(\cdot)$ is first chosen. The function $G(\cdot)$ is a smooth scalar value function of \mathbf{w} that measures the goodness-of-fit of the result of the estimation, and its derivative $g(\cdot)$, sometimes called an activation function, which typically appears in the algorithm as a nonlinear function.

- The first approach to ICA estimation is based on tensorial methods. The most well-known algorithms are the FOBI (for First-Order Blind Identification; see Cardoso, 1989, 1990) and JADE (for Joint Approximate Diagonalization of Eigenmatrices, see Cardoso and Souloumiac, 1993). Tensors are generalizations of linear operators - in particular, cumulant tensors are generalizations of the covariance matrix. The higher order decorrelation is approximate by the minimization of the higher order cumulants to solve the ICA model. Those methods are very burdensome in high dimensions (see Hyvärinen *et al.*, 2001).
- The second approach for the estimation of the ICA model is through maximum likelihood estimation. In maximum likelihood methods, one selects parameter values that give the highest probability score to the joint observations. If prior information on the parameters is taken into account, the method becomes a maximum *a posteriori* method. ICA algorithms based on the maximum likelihood method include the Bell-Sejnowski algorithm (see Bell and Sejnowski, 1995).
- The third approach is maximization of non-gaussianity of the components. According to the central limit theorem, sums of independent non-gaussian random variables are closer to being Gaussian than the original random variables. Thus a linear combination $\mathbf{y} = \sum \mathbf{b}_i \mathbf{x}_i$ of the observed variables \mathbf{x}_i (which, in turn, are linear combinations of the independent components \mathbf{s}_j) will be maximally non-gaussian if it equals one of the independent components \mathbf{s}_j . Thus, the task is to find \mathbf{w}_j such that the distribution of $\mathbf{y}_j = \mathbf{w}_j \mathbf{X}$ is as far from Gaussian as possible. In the literature, many measures of the Non-Gaussianity are proposed, often used is the higher order cumulants, such as *kurtosis* defined by:

$$Kurt(\mathbf{y}) = \frac{E(\mathbf{y}^4)}{[E(\mathbf{y}^2)]^2} \quad (2.7)$$

For a Gaussian variable, the *kurtosis* is equal to three. But this measure of non-gaussianity is not robust against outliers (Ruppert, 1987). Robust measures have been presented in Hyvärinen (1999). Non-gaussianity can be shown to have a rigorous connection to minimization of mutual information, expressed in terms of marginal entropies of the variables. The entropy of a random vector \mathbf{y} with density $f_{\mathbf{y}}(\mu)$ is:

$$H(\mathbf{y}) = - \int f_{\mathbf{y}}(\mu) \log f_{\mathbf{y}}(\mu) d\mu \quad (2.8)$$

The negentropy is defined as follow:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \quad (2.9)$$

Where \mathbf{y}_{gauss} is a Gaussian variable with the same correlation matrix as \mathbf{y} . Among all random variables of unit variance, a Gaussian variable has the largest entropy. Negentropy (equation 2.9) is a convenient measure of entropy; it is invariant for invertible linear transformations. It is always nonnegative, and zero for Gaussian variables.

To maximize the independence between random variables, one can make the variables as non-gaussian as possible. Since negentropy estimation requires the knowledge of the probability distribution function, this measure is difficult to estimate, and it is approximated in practice by order cumulants (see equation 2.10):

$$J(\mathbf{y}) = \frac{1}{12} E(\mathbf{y}^2)^2 + \frac{1}{48} kurt(\mathbf{y})^2 \quad (2.10)$$

An algorithm that can be used in all the previously listed ICA approaches is the FastICA algorithm (Hyvärinen, 1999; Hyvärinen *et al.*, 2001 and Hyvärinen and Oja, 1997) based

on the approximations of negentropy.

2.2.3 FastICA Algorithm

Although we can measure non-gaussianity by *kurtosis*, already we mentioned preview that this measure is however very sensitive to outliers. An alternative measure of non-gaussianity is negentropy, which is based on the information-theoretic quantity of entropy. A fundamental result of information theory is that a Gaussian variable has the largest entropy among all random variables with equal variance. Therefore, the negentropy defined above is always nonnegative, and it is zero for a Gaussian distribution. In FastICA algorithm, an approximated version of the negentropy is used such as:

$$J(\mathbf{y}) = \{E[G(\mathbf{y})] - E[G(N(0, 1))]\}^2 \quad (2.11)$$

where $N(0, 1)$ is a standard Gaussian random variable and $G(\cdot)$ is an even function. The even function $G(\cdot)$ is approximated by:

$$G(\mathbf{y}) = \frac{1}{\mathbf{a}} \log[\cosh(\mathbf{a}\mathbf{y})]$$

with $\mathbf{a} \in \mathbb{R}^*$.

Then the FastICA algorithm is an iterative fixed-point algorithm with the following update for \mathbf{W} :

$$\mathbf{W} \leftarrow E[\mathbf{X}g(\mathbf{W}\mathbf{X})] - E[g'(\mathbf{W}\mathbf{X})]\mathbf{W} \quad (2.12)$$

The nonlinear function $g(\cdot)$ is the derivative of the non-quadratic contrast function G , that measures negentropy or non-Gaussianity¹. A remarkable property of the FastICA algorithm is the high speed of convergence in the iterations.

¹If $G(\cdot)$ is an even function defined by equation (2.11) then, $g(\mathbf{y}) = G'(\mathbf{y}) = \tanh(\mathbf{a}\mathbf{y})$ and its derivative is given as: $g'(\mathbf{y}) = \mathbf{a}\{1 - [\tanh(\mathbf{a}\mathbf{y})]^2\}$.

2.2.4 An Example of ICA Estimation

We suppose that we have the following three hidden independent vectors,

$$s_1 = \cos\left(\frac{2\pi}{5}n\right)$$

with n a 10,000 linear integer vector between 0 and 10,000,

$$s_2 = 0.5 + 0.8 v$$

with $v \sim N(0, 1)$; and v is $(n \times 1)$ vector:

$$s_3 = 4 u$$

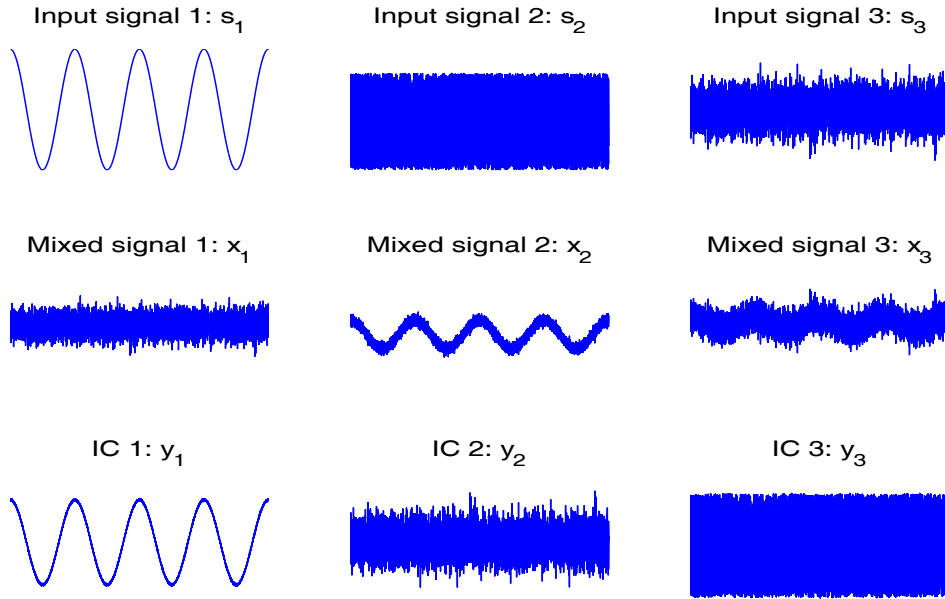
with $u \sim U(0, 1)$ and u is $(n \times 1)$ vector.

Figure 2.2 (top line) shows the simulated independent vectors \mathbf{S} . We use the following mixing matrix \mathbf{A} to compute the mixed matrix $\mathbf{X} = \mathbf{A} \mathbf{S}$ (Figure 2.2, middle line):

$$\mathbf{A} = \begin{pmatrix} 0.1 & 0.03 & 0.6 \\ 0.4 & -0.2 & 0.4 \\ 0.08 & 0.01 & -0.19 \end{pmatrix}$$

The question is: can ICA reveal the simulated independent vectors? The independent components computed with the procedure described in the FastICA algorithm (2.12) can be observed in figure 2.2 (bottom line). Note that only the previous mixing matrix (\mathbf{X}) is the input of the FastICA algorithm. The estimated ICs show clearly all the simulated independent vectors in the different order. ICA can therefore be used to extract the independent component (ICs) in the mixing process.

Figure 2.2. Illustration of Independent Component Analysis Extraction Technique



Source: Simulated data with the following parameters: $s_1 = \cos(2\pi 5^{-1}n)$, with n a 10,000 linear integer vector between 0 and 10,000; $s_2 = 0.5 + 0.8v$, with $v \sim N(0,1)$; and v is $(n \times 1)$ vector; $s_3 = 4u$, with $u \sim U(0,1)$ and u is $(n \times 1)$ vector. We use the following mixing matrix \mathbf{A} to compute the mixed components $\mathbf{X} = \mathbf{A} \mathbf{S}$: $\mathbf{A} = \begin{pmatrix} 0.1 & 0.03 & 0.6 \\ 0.4 & -0.2 & 0.4 \\ 0.08 & 0.01 & -0.19 \end{pmatrix}$.

The top line shows simulated independent vectors, the middle shows the mixed vectors and the bottom line, estimated independent vectors.

2.2.5 Reasons for Exploring ICA with Financial Time-series

The ICA technique assumes that a global signal caused by a set of signals is a superposition of elementary components that occur independently of each other. The technique is different from Principal Component Analysis (PCA) in the sense that it imposes - not only, as in the PCA, the second order zero correlation of components, but also higher order independences. In other words, Back and Weigend (1997) show that ICA dominates PCA in a financial context. In analyzing time-series with the ICA, we also need to select several dominant components that are useful when forecasting. The statistical

factor decomposition has to be validated with an economic interpretation. The hidden factors revealed in the ICA could include news (government intervention, natural disasters, political events), response to very large trades, or complex and still unexplained noise (Lai *et al.*, 1999; Moody and Yang, 1999; Cha and Chan, 2000; Cheung and Xu, 2001; Rojas and Moody, 2001 and Glasserman *et al.*, 2002).

2.3 Value-at-Risk Calculation

The VaR of a portfolio is the potential loss over a preset time horizon for a given probability (Duffee and Pan, 1997). In this section, we present the framework of VaR estimation in the case of multi-asset portfolio; we then present the link between the ICA and its actual computation and finally, we show some backtesting of VaR models.

2.3.1 General Framework of VaR Estimation

Let \mathbf{r}_i be the return of the i -th asset class of the entire portfolio during a specific time interval, also assume the risk factors associated with the entire portfolio be $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_K$. Since the return of an asset class is often caused by the change of the underlying risk factors², the return of the i -th asset class writes as³:

$$\mathbf{r}_i = \alpha_i + \sum_{k=1}^K \beta_{ik} \mathbf{F}_k + \epsilon_i \quad (2.13)$$

and,

$$\mathbf{r}_i - \alpha_i = B_i \mathbf{F}' + \epsilon_i$$

²Applications of factor models in finance include portfolio construction, sensitivity analysis. Besides, theories, such as Capital Asset Pricing Model (CAPM; Sharpe, 1964) and Arbitrage Pricing Theory (APT; Ross, 1976) or The Fama-French Three-factor Model (Fama and French, 1993), are built upon factor models.

³See Feuerverger and Wong (2000), Tsai (2004); for more details.

where B_i is a $(1 \times K)$ vector of constant coefficients for the i -th asset class and $\mathbf{F}(T \times K)$ is a random vector of risk factors. The random error ϵ_i with mean 0 accounts for the information not captured by the risk factors. Following equation (2.13), the mean of \mathbf{r}_i is $E(\mathbf{r}_i) = \alpha_i + \mathbf{B}_i[E(\mathbf{F})]'$ and the variance of \mathbf{r}_i is $Var(\mathbf{r}_i) = \mathbf{B}_i'\Sigma_{\mathbf{F}}\mathbf{B}_i + \sigma_{\epsilon}^2$ where $\Sigma_{\mathbf{F}}$ and σ_{ϵ}^2 are the covariance matrices of \mathbf{F} and ϵ_i , respectively. Note that the components β_{ik} of the coefficient vector \mathbf{B}_i will be zero if the i -th asset class is not exposed to the k -th risk factor.

If a portfolio consists in N asset classes, then the return of the portfolio can be written as:

$$\mathbf{R}_P = \sum_{i=1}^N \mathbf{w}_i \mathbf{r}_i = \mathbf{W}'\mathbf{R} \quad (2.14)$$

Combining equations (2.13) and (2.14), the return of the entire portfolio can be expressed as:

$$\mathbf{R}_P = \sum_{i=1}^N \mathbf{w}_i (\alpha_i + \sum_{k=1}^K \beta_{ik} \mathbf{F}_k + \epsilon_i) = \alpha_P + \sum_{i=1}^N \sum_{k=1}^K \mathbf{w}_i \beta_{ik} \mathbf{F}_k + \Xi_P \quad (2.15)$$

where $\alpha_P = \sum_{i=1}^N \mathbf{w}_i \alpha_i$ and Ξ_P is the aggregated random error term from all the asset classes of the portfolio.

An estimate of the VaR at the $\alpha\%$ threshold for the i -th asset of the portfolio is defined as:

$$VaR_{i,\alpha} = G_{i,\alpha}^{-1}(\mathbf{r}_i) \times \mathbf{V}_i \quad (2.16)$$

where $G_{i,\alpha}^{-1}(\cdot)$ is the quantile function of asset and \mathbf{V}_i is the market value of the i -th asset. Generally, $G_{i,\alpha}^{-1}(\cdot)$ is unknown and needs to be estimated from the existing data. If \mathbf{r}_i

follows a normal distribution, then:

$$VaR_{i,\alpha} = \Phi_{i,\alpha}^{-1}[\mathbf{r}_i^*] \times \mathbf{V}_i \quad (2.17)$$

where $\Phi_{i,\alpha}^{-1}(\cdot)$ is the quantile function of the normal distribution; $\mathbf{r}_i^* = [\mathbf{r}_i - E(\mathbf{r}_i)]\sigma_{\mathbf{r}_i}^{-2}$ is the centered reduced return on asset i ; $E(\mathbf{r}_i)$ and $\sigma_{\mathbf{r}_i}^2$ are respectively the mean and the variance of R_i .

Similarly, an estimate of the VaR at the $\alpha\%$ threshold for the entire portfolio can be defined as:

$$VaR_{P,\alpha} = G_{P,\alpha}^{-1}(\mathbf{R}_p) \times \mathbf{V}_P \quad (2.18)$$

where $G_{P,\alpha}^{-1}(\cdot)$ is the quantile function of the portfolio P whose the market value is \mathbf{V}_P . The VaR measure is not a coherent measure in the sense that it is not additive (see Artzner *et al.*, 1998 and Britten-Jones and Schaefer, 1999), meaning here that the Quantile Function is unknown and thus should be also estimated. Investors are in general not able to aggregate their portfolio because of liquidity problems and/or transaction costs between the different security markets (Artzner *et al.*, 1998).

2.3.2 Computing VaR with ICA

Since the return on an asset is caused by the change of the underlying risk factors, the return of the i -th asset writes as:

$$\mathbf{r}_i = \alpha_i + \sum_{k=1}^K \beta_{ik} \mathbf{F}_k + \epsilon_i \quad (2.19)$$

and:

$$\mathbf{r}_i - \alpha_i = \mathbf{B}_i \mathbf{F}' + \epsilon_i$$

where \mathbf{B}_i is a $(1 \times K)$ vector of constant sensitivities of the i -th asset to the $\mathbf{F}(T \times K)$ risk factors. The white noise ϵ_i accounts for the information not captured by the risk factors.

Note that the components β_{ik} of the coefficient vector \mathbf{B}_i will be zero if the i -th asset is not exposed to the k -th risk factor.

In traditional factor models, we require the factors to be uncorrelated to each other. It is possible that the factors are uncorrelated, but not independent. Therefore, the typical assumption on uncorrelated factors in the factor models cannot guarantee the return of the portfolio be free from the influence of other factors. On the contrary, if the factors are independent to each other, it is possible to construct such a portfolio. Loretan (1997) uses a PCA to extract uncorrelated factors in the context of VaR computation; Albanese *et al.* (2002) use delta-gamma approximation to reduce the dimension of the factor, whilst we propose here to use the ICA method for the extraction of Independent Factors (see Back and Trappenberg, 1999 and Lai *et al.*, 1999).

ICA extracts a set of systematic risk factors that are statistically independent. Following equation 2.19:

$$\mathbf{R} = \alpha + \mathbf{B}\mathbf{F} \iff \mathbf{F} = \mathbf{W}(\mathbf{R} - \alpha) \quad (2.20)$$

where \mathbf{R} is the $(N \times 1)$ vector of the asset returns, α is the $(N \times 1)$ vector of the asset expected returns, \mathbf{B} is the $(N \times K)$ factor loadings matrix, \mathbf{F} is the $(K \times 1)$ vector of the independent components and \mathbf{W} is a $(K \times N)$ unmixing matrix.

According to the central limit theorem, sums of independent random variables are Gaussian, so to maximize their statistical independence, ICA determines risk factors by maximizing their non-gaussianity, that is:

$$\mathbf{W}_k = \text{Arg Max } NG[\mathbf{W}_k(\mathbf{R} - \alpha)] \quad (2.21)$$

where \mathbf{W}_k is the k -th row vector $(1 \times N)$ of the unmixing matrix \mathbf{W} and $NG(\cdot)$ is a par-

ticular measure of the non-gaussianity of the k -th independent risk factor (such as for instance its *kurtosis* or its negentropy, see section 2).

We then fit a parametric model for the independent signals of ICA or uncorrelated factor of PCA. Since we need to capture the extreme events and focus on the tails of the portfolio return distribution (see Gonzalo and Olmo, 2004), a leptokurtic distribution, such as Pearson Type IV distribution or t Location Scale distribution are used⁴. The Gaussian distribution is also added to model the PCA factors and to compare with the ICA technique. We examine the accuracy of several VaR models by some statistical tests on forecast, of the return.

2.3.3 Backtesting VaR Models

The VaR Evaluation tests used in this chapter cover the Unconditional and Conditional Tests, the Dynamic Quantile Test, the Distributional Forecast Test and the Exception Magnitudes⁵. For the test, the exception indicator variable is defined as:

$$I_{Mt+1} = \begin{cases} 1 & \text{if } R_{pt+1} < VaR_{Mt}(\alpha) \\ 0 & \text{if } R_{pt+1} \geq VaR_{Mt}(\alpha) \end{cases} \quad (2.22)$$

where M is the model used to estimate VaR, t the current period, R_{pt+1} the actual portfolio on the period $t + 1$ and $VaR_{Mt}(\alpha)$ the one period VaR estimates from model M for the portfolio at the α quantile on period t .

Unconditional and Conditional Tests

The most used Backtest VaR models, based on the exception indicator, are the Unconditional and Conditional Tests. These VaR tests cover Unconditional Coverage Test

⁴This choice of leptokurtic models is motivated by the fact that ICA determines risk factors by maximizing their non-gaussianity (see previous section).

⁵In this section, we only give an summary of the test. For more details on these tests, see for instance Christoffersen (1998); Christoffersen and Pelletier (2003) or Ferreira and Lopez (2005).

(Kupiec, 1995), Independent Test and Conditional Coverage Test (Christoffersen, 1998). The conditional coverage test combines the Unconditional Coverage Test and the Independent Test.

Unconditional Coverage Test

The Unconditional test of the coverage of VaR estimates is a simply counts exceptions over the entire period. Kupiec (1995) shows that if the VaR model is well specified then the exception occurrences can be modeled as independent draws from a binomial distribution with a probability of occurrence equal to α percent. The likelihood ratio statistic is:

$$LR_{uc} = 2\{\log[\hat{\alpha}^x(1 - \hat{\alpha}^{T-x})] - \log[\alpha^x(1 - \alpha^{T-x})]\} \quad (2.23)$$

where x is the number of exceptions, T the number of observations and $\hat{\alpha} = x/T$ the unconditional coverage.

For the hypothesis that $\hat{\alpha} = \alpha$, the likelihood ratio LR_{uc} has an asymptotic $\chi^2(1)$ distribution.

Independent Test

This test detects the serial independence in the VaR model forecast. The Likelihood function for the test is:

$$LR_{ind} = 2[\log L_A - \log L_0] \quad (2.24)$$

which has an asymptotic $\chi^2(1)$ distribution (Christoffersen, 1998), where L_A is the likelihood function under the hypothesis of first-order Markov dependence:

$$L_A = (1 - \pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{10}} \pi_{11}^{T_{11}} \quad (2.25)$$

with T_{ij} denoting the number of observations in the state j after having been in i at the

previous period, $\pi_{01} = T_{01}/(T_{00} + T_{01})$ and $\pi_{11} = T_{11}/(T_{10} + T_{11})$. On the contrary, L_0 is the likelihood function under the hypothesis of independence ($\pi_{01} = \pi_{11}$):

$$L_0 = (1 - \pi)^{(T_{00}+T_{10})} \pi^{T\pi} \quad (2.26)$$

with $\pi = (T_{01} + T_{11})/T$.

Conditional Coverage Test

The Conditional test used here is a test of correct conditional coverage proposed by Christoffersen (1998). In this case, if the VaR estimated has correct conditional coverage, the exception variable must exhibit both correct unconditional coverage and serial independence. The Conditional coverage test (LR_{cc}) is therefore a joint test of these properties, and the relevant statistic is, with the previous notations:

$$LR_{cc} = LR_{uc} + LR_{ind} \quad (2.27)$$

which is asymptotically distributed as a $\chi^2(2)$.

Dynamic Quantile Test

The Dynamic Quantile (DQ) test is proposed by Engle and Manganelli (2004) to detect a good model for VaR estimation. The idea is that a good model should produce a sequence of unbiased and uncorrelated exception indicator variable. Under the null hypothesis, all coefficients as well as the intercept are zero in the regression (2.28) of the exception indicator variable on its past values and on current VaR estimate such as⁶:

$$I_{Mt+1} = \delta_0 + \sum_{k=1}^5 \delta_k I_{Mt-k+1} + \delta_6 VaR_{Mt} + \epsilon_{t+1} \quad (2.28)$$

⁶In this chapter, we use 5 lags following Ferreira and Lopez (2005).

were I_{Mt+1} is the exception indicator variable.

Distributional Forecast Test

This test is based on the fact that the VaR models are generally characterized by their distribution forecasts of portfolio returns; thus, the evaluations should be based directly on these forecasts. The object of interest in these evaluation methods is the observed quantile q_{Mt+1} , which is the quantile under the distribution forecast f_{Mt+1} in which the observed portfolio return R_{pt+1} actually falls. If the underlying VaR model is accurate, then its q_{Mt+1} series should be independent and uniformly distributed over the unit interval. The observed quantile q_{Mt+1} is defined as:

$$q_{Mt+1}(R_{pt+1}) = \int_{-\infty}^{R_{pt+1}} f_{Mt+1}(R_p) dR_p \quad (2.29)$$

To test the two properties of the series of observed quantiles (independence and uniform distribution), Diebold *et al.* (1998) propose to use the CUMSUM statistic, Christoffersen and Pelletier (2003) an Duration-based tests and Berkowitz (2001) proposes a Likelihood Ratio Test.

For the use of the two properties simultaneously, Berkowitz (2001) proposes the transformation $z_{Mt+1} = \Phi^{-1}(q_{Mt+1})$, which corresponds to the inverse of the standard normal cumulative distribution function of q_{Mt+1} . Under the null hypothesis, the LR_{dist} statistic is:

$$LR_{dist} = 2[L(\mu_M, \rho_M, \sigma_M^2) - L(0, 0, 1)] \quad (2.30)$$

which is an asymptotically distributed as a $\chi^2(3)$, with:

$$L(\mu_M, \rho_M, \sigma_M^2) = \sum_{t=1}^T \{-(1/2) \log[(2\pi\sigma_M^2)(1 - \rho_M^2)^{-1}] - L_z\}$$

$$-(1/2)(T - 1) \log(2\pi\sigma_M^2) - \sum_{t=1}^{T-1} (2\sigma_M^2)^{-1} (z_{Mt+1} - \mu_M - \rho_M z_{Mt})^2$$

with $L_z = \{[z_{Mt+1} - \mu_M(1 - \rho_M)^{-1}]^2\} [2\sigma_M^2 / (1 - \rho_M^2)]^{-1}$ and where (μ_M, ρ_M) are respect-

ively the conditional mean and $AR(1)$ coefficient corresponding to the z_{Mt+1} series, *i.e.* $z_{Mt+1} - \mu_M = \rho_M(z_{Mt} - \mu_M) + \epsilon_{t+1}$, ϵ_{t+1} being the normal random variable with zero mean and variance σ_M^2 .

Exception Magnitudes

The main idea of this test is that the magnitude of exceptions should be of primary interest to the various users of VaR models (see Hendricks, 1996; Berkowitz, 2001 and Ferreira and Lopez, 2005). For this test, as in the LR_{dist} test, Berkowitz (2001) transforms the empirical series into standard normal z_{Mt+1} series. The z_{Mt+1} values are then compared to the normal random variables with the desired coverage threshold of the VaR estimates. If the VaR model generating the empirical quantiles is correct, the γ_{Mt+1} series, defined by 2.31, should be identically distributed and (μ_M, σ_M) should equal $(.1)^7$.

$$\gamma_{Mt+1} = \begin{cases} z_{Mt+1} & \text{if } z_{Mt+1} < \Phi^{-1}(\alpha) \\ 0 & \text{if } z_{Mt+1} \geq \Phi^{-1}(\alpha) \end{cases} \quad (2.31)$$

where Φ is the standard normal cumulative distribution function. Finally, the corresponding test statistic is:

$$LR_{mag} = 2[L_{mag}(\mu_M, \sigma_M) - L(0, 1)] \quad (2.32)$$

which is asymptotically distributed $\chi^2(2)$, where:

$$L_{mag}(\mu_M, \sigma_M) = \sum_{\gamma_{Mt+1}=0} \log\{1 - \Phi[(\Phi^{-1}(\alpha) - \mu_M)\sigma_M^{-1}]\} \\ + \sum_{\gamma_{Mt+1} \neq 0} \{-(1/2) \log(2\pi\sigma_M^2) - (\gamma_{Mt+1} - \mu_M)^2(2\sigma_M^2)^{-1} - \log[\Phi((\Phi^{-1}(\alpha) - \mu_M)\sigma_M^{-1})]\}.$$

2.4 Experiments and Results

To investigate the effectiveness of ICA techniques for Value-at-Risk computation, we apply ICA to data from the French Stock Market. We use intraday prices from January

⁷ μ_M and σ_M are respectively the unconditional mean and standard deviation of the z_{Mt+1} series.

2002 until December 2006 of 43 firms. The preprocessing consists of two steps: we obtain the intraday stock returns (in this chapter, 10 and 15 minutes subsampling are used) and are normalize the resulting values to lie within the range $[-1,1]$.

We performed ICA on the stock returns using the FastICA algorithm (Hyvärinen *et al.*, 2001) described in previous section. In all the experiments, we assume that the number of stocks equals the number of sources supplied to the mixing model. Figure 2.3 (a) shows an Independent Components (IC) obtained from the algorithm and figure 2.3 (b) shows the PCA factors; and the corresponding statistics are given in table 2.1 and table 2.2.

Table 2.1. Summary Statistics of Independent Factors from ICA

	Min.	Max.	Mean	Stand. Dev.	Median	Skewness	<i>kurtosis</i>	Negative Return
10 minutes								
IC1	-7.95	6.40	0.00	1.00	-0.02	0.01	7.73	51%
IC2	-13.34	7.45	0.00	1.00	0.02	-0.50	14.09	49%
IC3	-5.55	16.39	0.00	1.00	-0.01	0.87	18.77	51%
IC4	-6.41	5.83	0.01	1.00	0.03	-0.30	6.20	48%
IC5	-8.36	7.37	0.01	1.00	0.01	-0.63	10.09	49%
IC6	-11.87	14.64	0.00	1.00	-0.01	1.16	27.30	50%
IC7	-10.25	21.06	0.00	1.00	-0.01	-1.44	44.70	50%
15 minutes								
IC1	-7.52	7.15	0.01	1.00	0.02	-0.26	8.26	49%
IC2	-15.58	8.03	0.00	1.00	-0.01	-0.72	18.74	50%
IC3	-10.46	5.06	0.00	1.00	0.00	-0.26	7.67	50%
IC4	-8.87	8.58	0.00	1.00	0.02	-0.35	11.46	49%
IC5	-17.95	8.86	-0.01	1.00	-0.02	-0.64	28.01	51%
IC6	-10.93	12.92	-0.01	1.00	-0.02	0.82	20.11	51%
IC7	-15.87	7.99	0.00	1.00	0.01	-0.56	19.49	49%

Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. This table reports summary statistics on independent factors of the ICA extraction technique.

From the original and the rebuilt stock prices, we compute a VaR and the ICA-based VaR. Figures 2.4 to 2.7 represent the returns of an equally weighted portfolio and the corresponding alternative VaRs: namely, Normal-VaR (assumes the normality distribution of the portfolio returns), Historical-VaR (approximates portfolio profit/loss distribution based on historical percentile), RiskMetrics-VaR (Value-at-Risk method based on the

Table 2.2. Summary Statistics of Factors from PCA

	Min.	Max.	Mean	Stand. Dev.	Median	Skewness	<i>kurtosis</i>	Negative Return
10 minutes								
PC1	-0.13	0.06	0.00	0.01	0.00	-1.12	32.68	50%
PC2	-0.05	0.05	0.00	0.01	0.00	0.06	12.81	50%
PC3	-0.06	0.03	0.00	0.01	0.00	-0.69	10.60	49%
PC4	-0.05	0.03	0.00	0.00	0.00	-0.33	11.42	49%
PC5	-0.03	0.02	0.00	0.00	0.00	-0.22	7.28	50%
PC6	-0.02	0.03	0.00	0.00	0.00	0.40	7.24	51%
PC7	-0.03	0.02	0.00	0.00	0.00	-0.62	11.09	49%
15 minutes								
PC1	-0.06	0.11	0.00	0.01	0.00	0.67	18.49	50%
PC2	-0.07	0.06	0.00	0.01	0.00	-0.37	15.22	50%
PC3	-0.04	0.04	0.00	0.01	0.00	-0.33	8.17	49%
PC4	-0.05	0.03	0.00	0.00	0.00	-0.32	12.16	50%
PC5	-0.04	0.06	0.00	0.00	0.00	0.52	18.55	50%
PC6	-0.03	0.03	0.00	0.00	0.00	0.38	8.46	50%
PC7	-0.02	0.04	0.00	0.00	0.00	0.24	9.47	50%

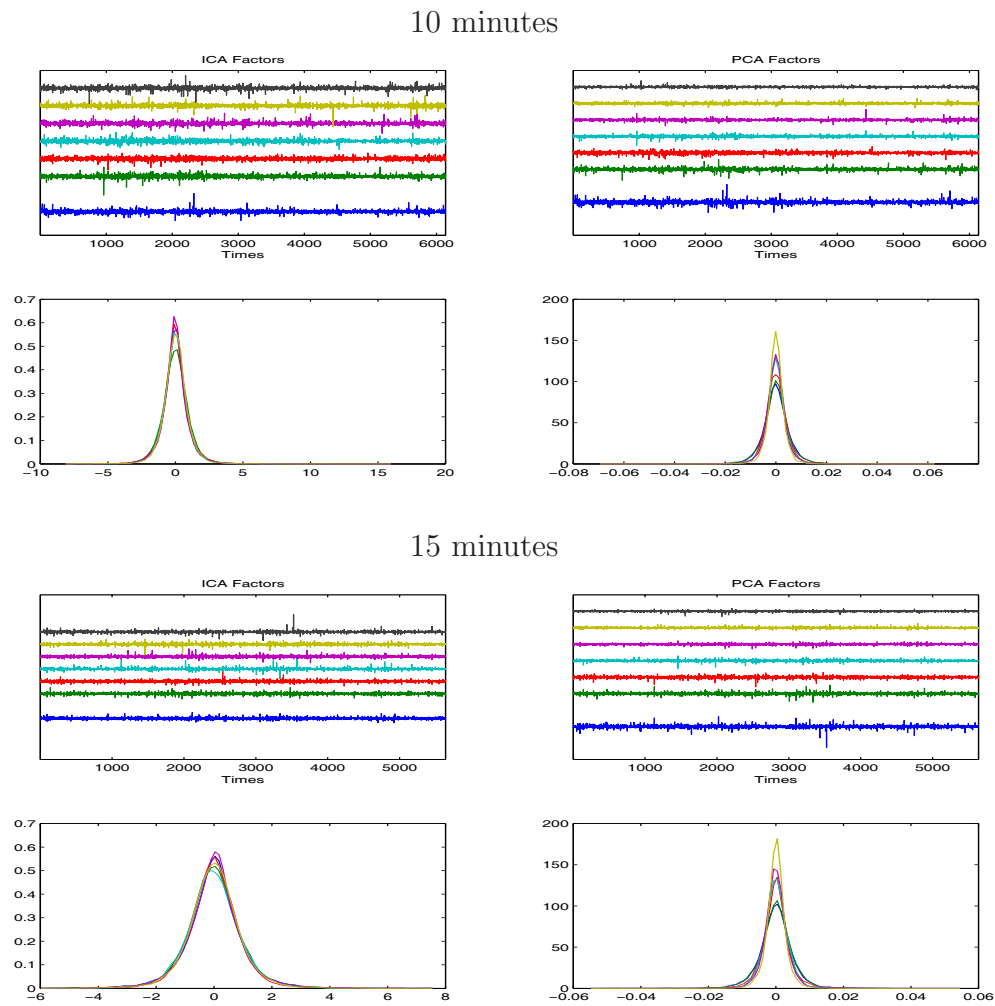
Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. This table reports summary statistics on PCA factors.

RiskMetrics correlation matrix) and Student-VaR (assumes that the portfolio returns is generated by a Student distribution)⁸. ICA-based VaR consists in computing these previous VaR models with the ICA reconstructed stock prices. The VaR evaluation is then performed using Loss Function based Backtests (see Christoffersen and Pelletier, 2003 and Campbell, 2007)⁹.

⁸See Giot and Laurent (2004).

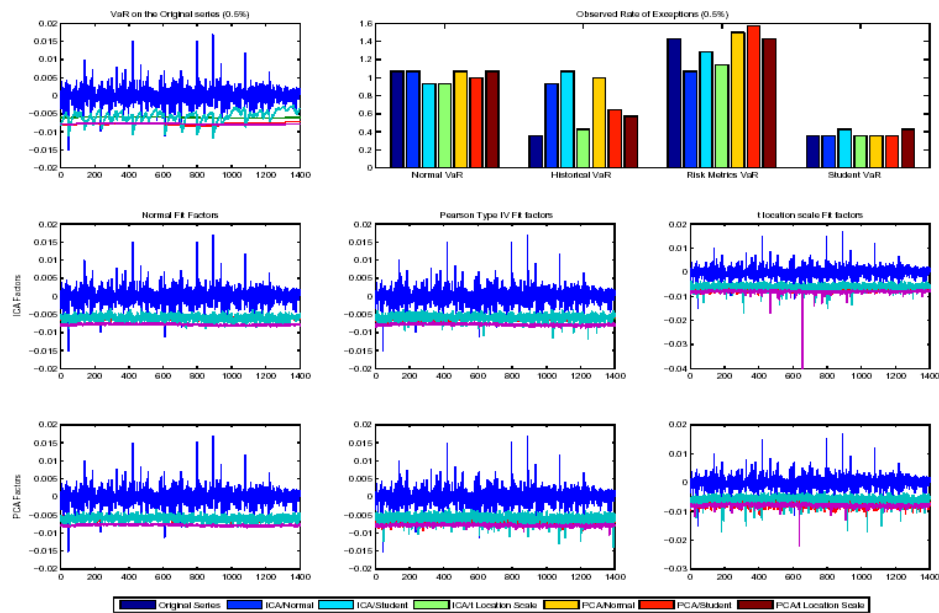
⁹Another multivariate form of the test exists in the literature with some probability $(1-\alpha)$ in the VaR violations sequence (Hurlin and Tokpavi, 2007).

Figure 2.3. Independent and Principal Components with corresponding Densities for 10 and 15-minute quotes

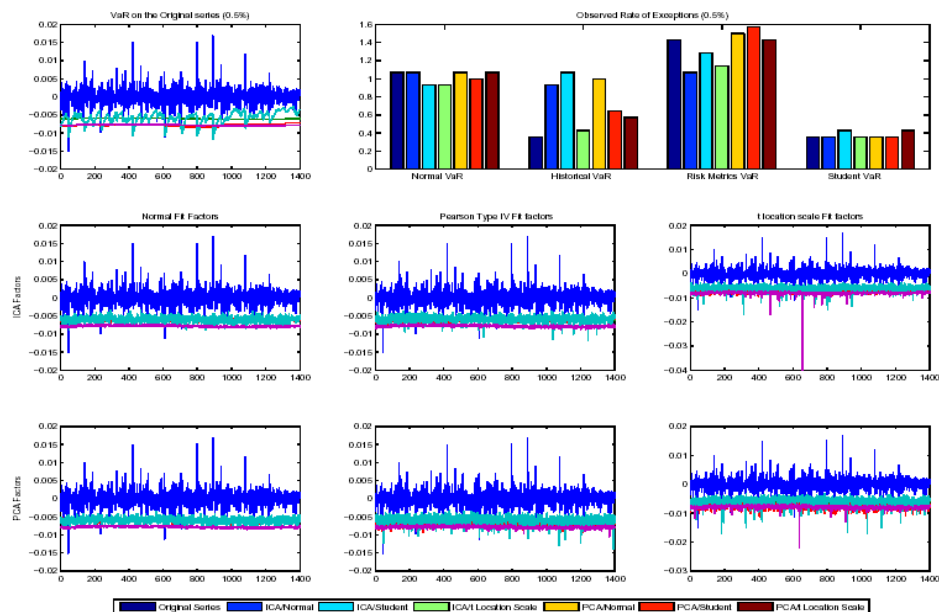


Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. (a) Estimated Independent Vectors (top) and corresponding Density (bottom), (b) Estimated PCA Factors (top) and corresponding Densities (bottom)

Figure 2.4. A VaR Forecasts Comparison at 95% confidence threshold
10 minutes

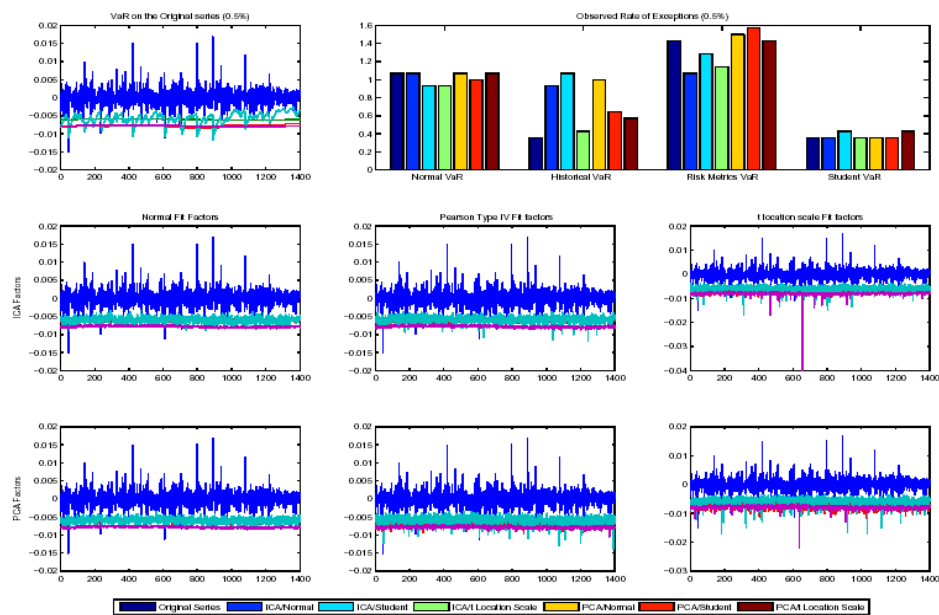


15 minutes

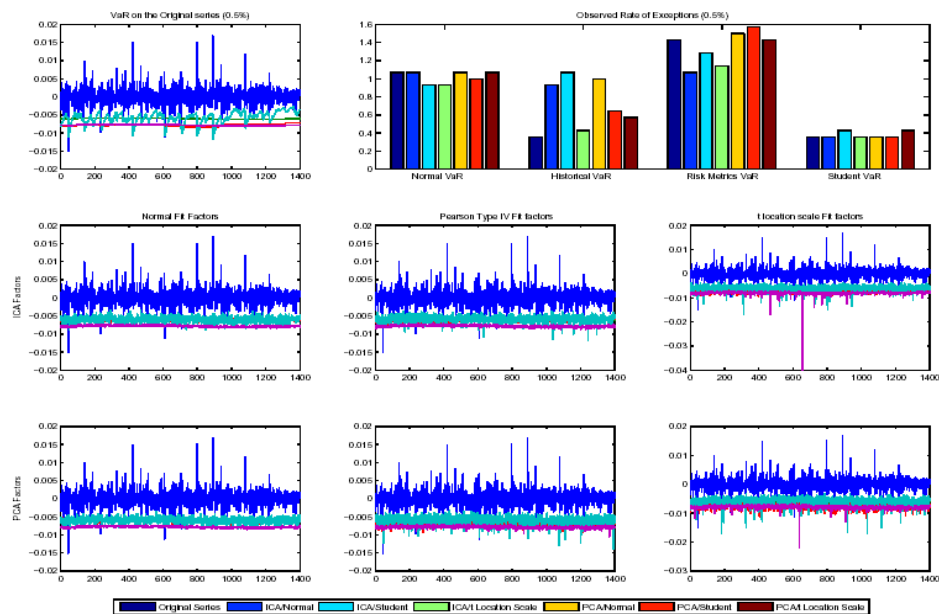


Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006.

Figure 2.5. A VaR Forecasts Comparison at 97.5% confidence threshold
10 minutes

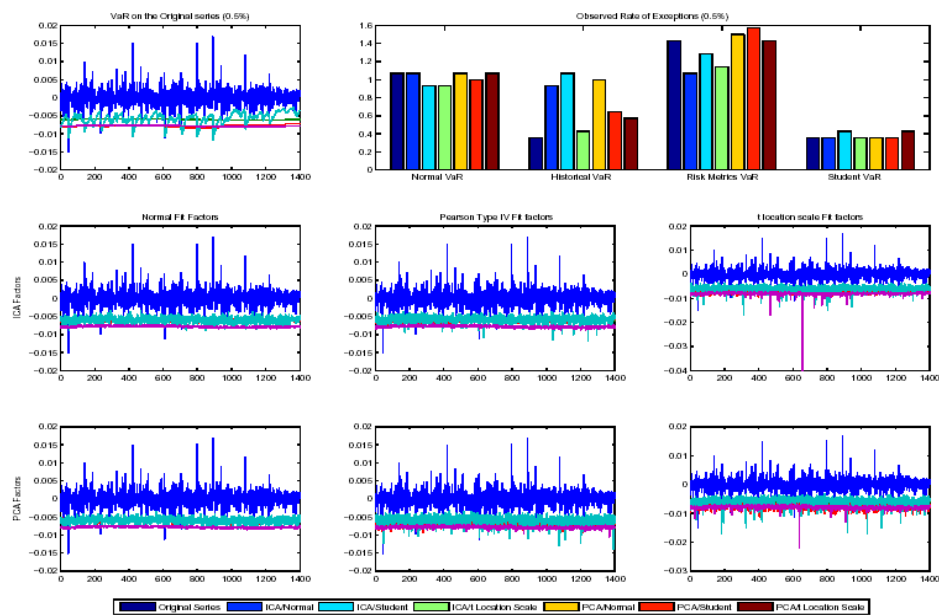


15 minutes

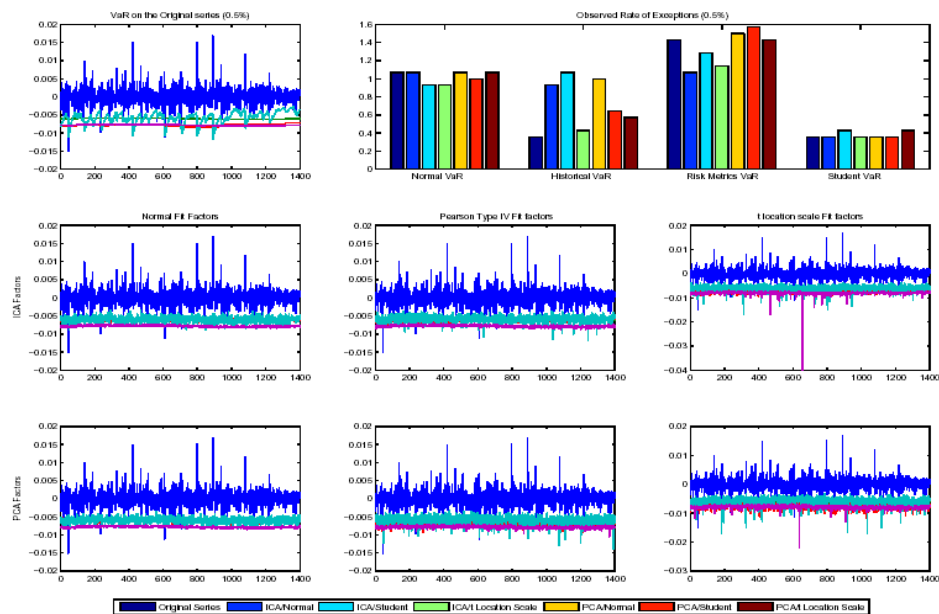


Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006.

Figure 2.6. A VaR Forecasts Comparison at 99% confidence threshold 10 minutes

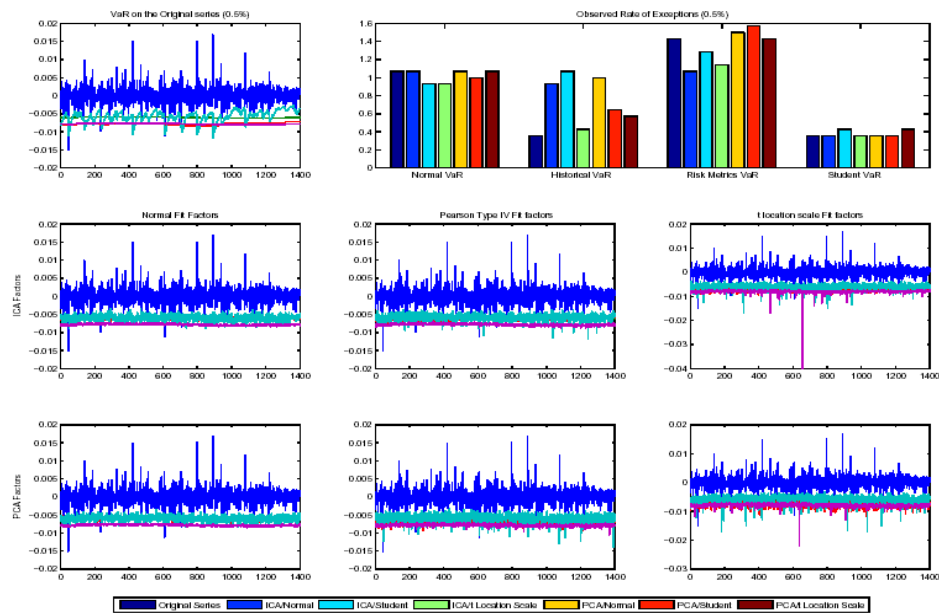


15 minutes

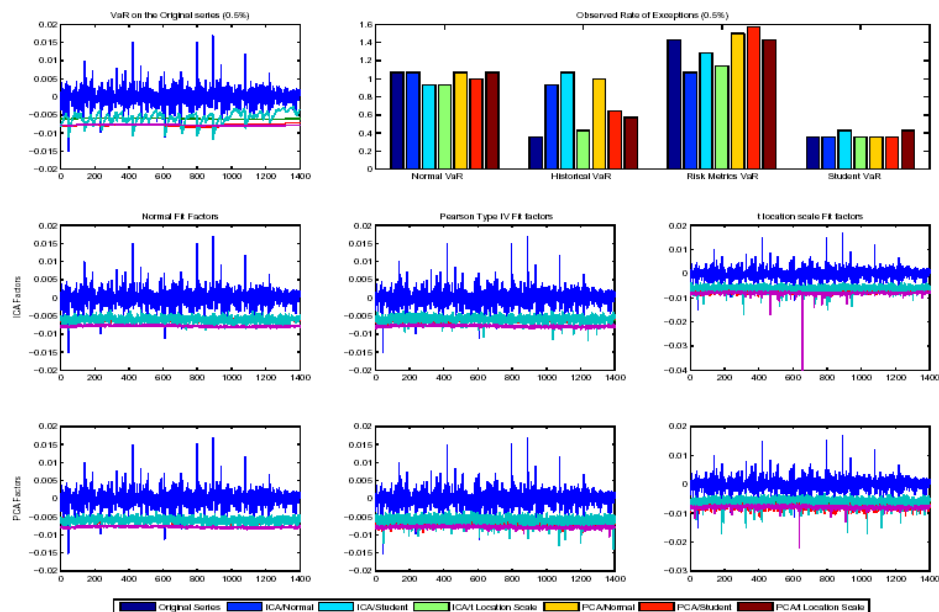


Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006.

Figure 2.7. A VaR Forecasts Comparison at 99.5% confidence threshold
10 minutes



15 minutes



Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006.

Table 2.3. Value-at-Risk Measures of Portfolio Returns

	Normal Distribution				Historical				Risk Metrics				<i>t</i> -distribution			
	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
10 minutes																
Original Series	0.38	0.46	0.54	0.60	0.35	0.45	0.61	0.78	0.22	0.26	0.31	0.35	0.43	0.54	0.67	0.78
ICA with Normal	0.38	0.46	0.54	0.60	0.39	0.47	0.56	0.63	0.38	0.46	0.54	0.60	0.43	0.54	0.68	0.79
ICA with Pearson Type IV	0.39	0.46	0.55	0.61	0.37	0.44	0.54	0.62	0.40	0.48	0.57	0.63	0.44	0.54	0.68	0.79
ICA with <i>t</i> -Location Scale	0.40	0.48	0.57	0.63	0.39	0.48	0.58	0.68	0.37	0.44	0.53	0.58	0.45	0.56	0.71	0.82
PCA with Normal	0.39	0.46	0.55	0.61	0.38	0.46	0.54	0.60	0.39	0.46	0.55	0.61	0.44	0.54	0.68	0.79
PCA with Pearson Type IV	0.38	0.45	0.53	0.59	0.36	0.48	0.60	0.67	0.37	0.44	0.53	0.58	0.43	0.53	0.67	0.77
PCA with <i>t</i> -Location Scale	0.38	0.45	0.54	0.59	0.36	0.46	0.65	0.74	0.25	0.30	0.36	0.40	0.43	0.53	0.67	0.77
15 minutes																
Original Series	0.48	0.58	0.68	0.76	0.41	0.55	0.73	0.96	0.40	0.47	0.56	0.62	0.55	0.68	0.85	0.99
ICA with Normal	0.48	0.57	0.68	0.75	0.48	0.56	0.69	0.75	0.47	0.56	0.67	0.74	0.54	0.67	0.84	0.98
ICA with Pearson Type IV	0.49	0.58	0.69	0.77	0.50	0.65	0.89	1.05	0.52	0.62	0.74	0.82	0.55	0.69	0.87	1.00
ICA with <i>t</i> -Location Scale	0.47	0.56	0.66	0.73	0.45	0.57	0.69	0.79	0.43	0.51	0.60	0.67	0.53	0.66	0.83	0.96
PCA with Normal	0.48	0.57	0.68	0.75	0.48	0.56	0.68	0.79	0.43	0.51	0.60	0.67	0.54	0.67	0.85	0.98
PCA with Pearson Type IV	0.46	0.56	0.66	0.74	0.50	0.61	0.87	1.08	0.37	0.44	0.53	0.59	0.53	0.66	0.83	0.96
PCA with <i>t</i> -Location Scale	0.50	0.60	0.71	0.78	0.43	0.56	0.79	0.98	0.51	0.60	0.71	0.79	0.57	0.70	0.88	1.02

Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. This table shows mean VaR estimates of an equally weighted portfolio of stocks in percentage for the alternative VaR models under the 5%, 2.5%, 1%, 0.5% significance thresholds.

Table 2.4. Conditional Value-at-Risk (CVaR) of an Equally Weighted Portfolio for different VaR-based Measures

	Normal Distribution				Historical				Risk Metrics				t -distribution			
	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
10 minutes																
Original Series	0.02	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.05	0.04	0.03	0.03	0.02	0.01	0.01	0.01
ICA with Normal	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.00
ICA with Pearson Type IV	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.00
ICA with t Location Scale	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.00
PCA with Normal	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.00
PCA with Pearson Type IV	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.01
PCA with t Location Scale	0.02	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01
15 minutes																
Original Series	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.04	0.03	0.02	0.01	0.02	0.01	0.01	0.01
ICA with Normal	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.02	0.01	0.01	0.01
ICA with Pearson Type IV	0.03	0.02	0.01	0.01	0.03	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.01
ICA with t Location Scale	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0.01
PCA with Normal	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0.01
PCA with Pearson Type IV	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.04	0.03	0.02	0.02	0.02	0.01	0.01	0.01
PCA with t Location Scale	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0.01	0.02	0.01	0.01	0.01

Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. This table shows Conditional Value-at-Risk estimates of an equally weighted portfolio of stocks in percentage for the alternative VaR models under the 5%, 2.5%, 1%, 0.5% significance thresholds.

The failure rates are given in Tables 2.5 to 2.9. Generally speaking, the results are quite good for all VaR computation methods. But we observe that ICA-VaR with Pearson Type IV distribution has many best scores at high confidence threshold (1% and 0.5%). This result confirms that the classical methods of VaR evaluation will be improved by using High-frequency data.

2.5 Conclusion

In this chapter, we have applied the Independent Component Analysis (ICA) to compute the Value-at-Risk of a portfolio. We show, in an environment characterized by non-linearities and non-Gaussianity, that the ICA methodology can advantageously reveal an underlying structure in intraday data that can be exploited for risk measurement.

We have also show with different VaR specifications (Normal, Historical, Risk-Metrics and Student) and backtests (Conditional and unconditional Tests), that ICA-VaR is well defined for the intra-daily data. Interestingly, the “noise” that we observed in the intra-day data may be attributed to signals with a small relative range. The VaR computation with the denoised data (ICA-VaR) is therefore more robust than VaR computation done directly with the original data or with the PCA factors extraction.

Finally, based on our first results, the ICA technique, directly dealing with the factors that drive the market, could have an interest in some further financial applications, such as stress test assessment or asset pricing. The use of the Independent Component Analysis for the extraction of the risk factors in general offers new tracks for factor model and the evaluation of measurements of risk in particular. The results obtained on the calculation of the VaR in high frequency context show that this technique is an alternative to Principal Components Analysis whenever the Gaussian assumption of distribution cannot be proved. Some extensions can be made. The first extension comes from the

modelling of the risk factors determined by the algorithm. We can test families of distributions integrating higher moments in particular starting from the models Skew-student (Keith and Shen, 2004 and Adcock and Meade, 2007).

Another extension can be to test other approaches of extractions of the independent factors as the use of the Copula Independent Component Analysis (COPICA, Chen *et al.*, 2008) or the methods with constraints. Indeed, the ICA model presented in this chapter does not allow, like in the case of PCA, to reduce the number of factors; it determines the number of statistically independent factors automatically. We can start from an Analysis in Independent Components with constraint Independent Component Analysis (cICA, Rajapakse, 2000 and 2005), to reduce the dimension of the data while keeping the character of independence of the factors. Moreover, we can suppose that the data observed are combinations of the independent sources which are affected by noises. For this purpose, Chang and Zhang (2006) and Kopriva and Seršić (2008) proposed algorithms which make it possible to filter the noises by Wavelet method, before determining the independent components.

Table 2.5. Observed Frequency of Exceptions

	Normal Distribution				Historical				Risk Metrics				t -distribution			
	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
10 minutes																
Original Series	3.6	2.4	1.4	1.1	5.3	2.5	0.9	0.4	5.4	3.7	2.2	1.4	2.9	1.4	0.7	0.4
ICA with Normal	3.7	2.5	1.4	1.1	3.7	2.4	1.4	0.9	4.5	2.9	1.6	1.1	2.7	1.4	0.6	0.4
ICA with Pearson Type IV	3.8	2.5	1.4	0.9	4.9	2.4	1.1	0.4	4.4	3.3	1.9	1.3	3.0	1.4	0.7	0.5
ICA with t -Location Scale	3.8	2.4	1.4	0.9	3.9	2.4	0.9	0.4	4.6	2.9	1.6	1.1	2.7	1.5	0.6	0.4
PCA with Normal	3.7	2.4	1.5	1.1	3.9	2.5	1.4	1.0	4.0	2.9	1.9	1.5	2.9	1.5	0.7	0.4
PCA with Pearson Type IV	3.7	2.5	1.4	1.0	4.5	2.7	1.4	0.6	4.5	2.8	1.9	1.6	2.8	1.5	0.6	0.4
PCA with t -Location Scale	3.6	2.6	1.5	1.1	4.8	2.4	0.9	0.6	4.9	3.6	1.8	1.4	2.9	1.6	0.6	0.4
15 minutes																
Original Series	3.2	1.9	1.4	1.0	4.1	2.0	1.1	0.7	4.9	3.4	2.2	1.9	2.1	1.4	0.9	0.4
ICA with Normal	3.1	1.9	1.4	1.0	3.4	1.9	1.5	0.9	3.2	2.1	1.4	0.9	2.1	1.4	0.9	0.5
ICA with Pearson Type IV	3.1	2.1	1.4	0.9	4.6	2.3	1.1	0.7	3.6	2.4	1.8	1.5	2.9	1.4	0.8	0.5
ICA with t -Location Scale	3.3	1.8	1.4	1.1	3.6	2.1	1.1	0.7	3.9	2.6	1.4	1.3	2.4	1.4	0.8	0.5
PCA with Normal	3.1	1.9	1.4	1.2	3.0	2.0	1.4	1.1	3.1	2.4	1.6	1.1	2.4	1.4	0.8	0.4
PCA with Pearson Type IV	3.0	2.0	1.6	1.0	3.6	1.9	1.1	0.7	3.6	2.1	1.5	1.2	2.1	1.6	0.9	0.4
PCA with t -Location Scale	3.4	2.0	1.5	1.1	3.9	2.0	0.9	0.7	4.0	2.6	1.7	1.5	2.3	1.5	0.9	0.6

Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. This table presents the observed frequency in percentage of exceptions for the alternative VaR models under the 5%, 2.5%, 1%, 0.5% significance thresholds. Note that, an exception is defined by the indicator variable I_{Mt+1} for the given model M . *i.e.* $I_{Mt+1} = 1_{\{R_{pt+1} - VaR_{Mt}(\alpha) < 0\}}$, where R_{pt+1} is the portfolio return on time $t+1$ and $VaR_{Mt}(\alpha)$ is the Value-at-Risk calculated on the previous time with significance threshold α .

Table 2.6. Unconditional Coverage Test Results

	Normal Distribution				Historical				Risk Metrics				t -distribution			
	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
10 minutes																
Original Series	0.01	0.86	0.13	0.01	0.90	1.00	0.79	0.42	0.54	0.01	0.00	0.00	0.00	0.01	0.26	0.42
ICA with Normal	0.02	1.00	0.13	0.01	0.02	0.73	0.13	0.04	0.38	0.40	0.05	0.01	0.00	0.01	0.15	0.42
ICA with Pearson Type IV	0.03	1.00	0.13	0.04	0.62	0.50	0.90	0.70	0.32	0.07	0.00	0.00	0.00	0.01	0.26	0.70
ICA with t -Location Scale	0.03	0.86	0.13	0.04	0.06	0.86	0.79	0.70	0.46	0.40	0.03	0.00	0.00	0.01	0.15	0.42
PCA with Normal	0.02	0.86	0.08	0.01	0.04	1.00	0.13	0.02	0.08	0.40	0.00	0.00	0.00	0.01	0.26	0.42
PCA with Pearson Type IV	0.02	1.00	0.13	0.02	0.38	0.61	0.13	0.47	0.38	0.50	0.00	0.00	0.00	0.01	0.15	0.42
PCA with t -Location Scale	0.01	0.73	0.08	0.01	0.71	0.73	0.79	0.71	0.81	0.02	0.01	0.00	0.00	0.02	0.15	0.70
15 minutes																
Original Series	0.00	0.15	0.20	0.02	0.13	0.21	0.79	0.29	0.90	0.05	0.00	0.00	0.00	0.01	0.58	0.70
ICA with Normal	0.00	0.15	0.13	0.02	0.00	0.15	0.08	0.04	0.00	0.29	0.13	0.04	0.00	0.01	0.58	1.00
ICA with Pearson Type IV	0.00	0.29	0.13	0.02	0.01	0.21	0.30	0.09	0.91	0.73	0.01	0.00	0.00	0.01	0.58	1.00
ICA with t -Location Scale	0.00	0.07	0.13	0.00	0.01	0.38	0.60	0.29	0.04	0.73	0.20	0.00	0.00	0.01	0.40	1.00
PCA with Normal	0.00	0.15	0.20	0.00	0.00	0.21	0.13	0.00	0.00	0.86	0.05	0.00	0.00	0.01	0.40	0.70
PCA with Pearson Type IV	0.00	0.21	0.05	0.02	0.01	0.11	0.60	0.29	0.01	0.29	0.08	0.00	0.00	0.02	0.58	0.70
PCA with t -Location Scale	0.00	0.21	0.08	0.00	0.06	0.21	0.79	0.29	0.08	0.73	0.01	0.00	0.00	0.01	0.58	0.71

Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. This table shows the p-value for LR_{uc} statistics for the alternative VaR models under the 5%, 2.5%, 1%, 0.5% significance thresholds. The LR_{uc} statistics are asymptotically distributed $\chi^2(1)$; the cells in bold font indicate rejection of null hypothesis of correct VaR estimates at the 5% significance threshold.

Table 2.7. Independent Test Results

	Normal Distribution				Historical				Risk Metrics				<i>t</i> -distribution			
	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
10 minutes																
Original Series	0.48	0.19	0.45	0.57	0.09	0.89	0.62	0.85	0.62	0.96	0.24	0.45	0.89	0.45	0.70	0.85
ICA with Normal	0.45	0.18	0.45	0.57	0.45	0.21	0.45	0.62	0.58	0.89	0.40	0.57	0.15	0.45	0.73	0.85
ICA with Pearson Type IV	0.42	0.18	0.45	0.62	0.15	0.13	0.38	0.57	0.21	0.08	0.32	0.49	0.93	0.45	0.70	0.82
ICA with <i>t</i> -Location Scale	0.42	0.19	0.45	0.62	0.36	0.19	0.62	0.82	0.96	0.89	0.38	0.54	0.97	0.42	0.73	0.85
PCA with Normal	0.45	0.19	0.42	0.57	0.39	0.18	0.45	0.59	0.03	0.12	0.30	0.42	0.89	0.42	0.70	0.85
PCA with Pearson Type IV	0.45	0.18	0.45	0.59	0.19	0.97	0.45	0.73	0.58	0.93	0.55	0.36	0.93	0.42	0.73	0.85
PCA with <i>t</i> -Location Scale	0.51	0.98	0.42	0.57	0.13	0.81	0.62	0.76	0.42	0.51	0.47	0.45	0.85	0.40	0.73	0.82
15 minutes																
Original Series	0.68	0.30	0.47	0.59	0.78	0.28	0.57	0.70	0.74	0.07	0.24	0.32	0.27	0.45	0.65	0.82
ICA with Normal	0.76	0.30	0.45	0.59	0.61	0.30	0.42	0.62	0.68	0.27	0.45	0.62	0.25	0.45	0.65	0.79
ICA with Pearson Type IV	0.09	0.27	0.45	0.59	0.51	0.28	0.49	0.65	0.05	0.21	0.34	0.42	0.21	0.45	0.65	0.79
ICA with <i>t</i> -Location Scale	0.65	0.34	0.45	0.54	0.48	0.25	0.54	0.70	0.04	0.16	0.47	0.49	0.19	0.45	0.68	0.79
PCA with Normal	0.09	0.30	0.47	0.52	0.11	0.28	0.45	0.54	0.10	0.19	0.40	0.54	0.21	0.45	0.68	0.82
PCA with Pearson Type IV	0.11	0.28	0.40	0.59	0.48	0.32	0.54	0.70	0.48	0.27	0.42	0.52	0.25	0.40	0.65	0.82
PCA with <i>t</i> -Location Scale	0.58	0.28	0.42	0.54	0.36	0.28	0.62	0.70	0.86	0.98	0.36	0.42	0.22	0.42	0.65	0.76

Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. This table shows the p-value for LR_{ind} statistics for the alternative VaR models under the 5%, 2.5%, 1%, 0.5% significance thresholds. The LR_{ind} statistics are asymptotically distributed $\chi^2(1)$; the cells in bold font indicate rejection of null hypothesis of correct VaR estimates at the 5% significance threshold.

Table 2.8. Conditional Coverage Test Results

	Normal Distribution				Historical				Risk Metrics				<i>t</i> -distribution			
	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
10 minutes																
Original Series	0.04	0.42	0.24	0.03	0.24	0.99	0.85	0.71	0.73	0.02	0.00	0.00	0.00	0.02	0.49	0.71
ICA with Normal	0.05	0.41	0.24	0.03	0.05	0.42	0.24	0.11	0.59	0.70	0.10	0.03	0.00	0.02	0.34	0.71
ICA with Pearson Type IV	0.07	0.41	0.24	0.11	0.31	0.26	0.06	0.03	0.28	0.04	0.01	0.00	0.00	0.02	0.49	0.90
ICA with <i>t</i> -Location Scale	0.07	0.42	0.24	0.11	0.11	0.42	0.85	0.90	0.76	0.70	0.06	0.01	0.00	0.03	0.34	0.71
PCA with Normal	0.05	0.42	0.16	0.03	0.09	0.41	0.24	0.06	0.02	0.22	0.00	0.00	0.00	0.03	0.49	0.71
PCA with Pearson Type IV	0.05	0.41	0.24	0.06	0.29	0.88	0.24	0.73	0.59	0.79	0.01	0.00	0.00	0.03	0.34	0.71
PCA with <i>t</i> -Location Scale	0.03	0.94	0.16	0.03	0.30	0.91	0.85	0.89	0.70	0.04	0.02	0.00	0.00	0.04	0.34	0.90
15 minutes																
Original Series	0.00	0.21	0.34	0.06	0.31	0.26	0.82	0.53	0.94	0.03	0.00	0.00	0.00	0.02	0.77	0.90
ICA with Normal	0.00	0.21	0.24	0.06	0.01	0.21	0.16	0.11	0.00	0.31	0.24	0.11	0.00	0.02	0.77	0.97
ICA with Pearson Type IV	0.00	0.31	0.24	0.06	0.03	0.26	0.47	0.21	0.01	0.42	0.02	0.00	0.00	0.02	0.77	0.97
ICA with <i>t</i> -Location Scale	0.01	0.12	0.24	0.01	0.04	0.35	0.72	0.53	0.01	0.35	0.34	0.00	0.00	0.02	0.65	0.97
PCA with Normal	0.00	0.21	0.34	0.00	0.00	0.26	0.24	0.01	0.00	0.42	0.10	0.01	0.00	0.02	0.65	0.90
PCA with Pearson Type IV	0.00	0.26	0.10	0.06	0.04	0.17	0.72	0.53	0.04	0.31	0.16	0.00	0.00	0.04	0.77	0.90
PCA with <i>t</i> -Location Scale	0.01	0.26	0.16	0.01	0.11	0.26	0.85	0.53	0.20	0.94	0.03	0.00	0.00	0.03	0.77	0.89

Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. This table shows the p-value for LR_{cc} statistics for the alternative VaR models under the 5%, 2.5%, 1%, 0.5% significance thresholds. The LR_{cc} statistics are asymptotically distributed $\chi^2(2)$; the cells in bold font indicate rejection of null hypothesis of correct VaR estimates at the 5% significance threshold.

Table 2.9. Dynamic Quantile Test Results

	Normal Distribution				Historical				Risk Metrics				<i>t</i> -distribution			
	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
10 minutes																
Original Series	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ICA with Normal	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.59	0.00	0.00	0.00	0.00
ICA with Pearson Type IV	0.00	0.15	0.00	0.43	0.06	0.10	0.22	0.40	0.00	0.00	0.99	0.90	0.00	0.00	0.00	0.00
ICA with <i>t</i> -Location Scale	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00
PCA with Normal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PCA with Pearson Type IV	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.00	0.00	0.00	0.00
PCA with <i>t</i> -Location Scale	0.06	1.00	1.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.99	1.00	0.00	0.00	0.00	0.00
15 minutes																
Original Series	0.00	0.00	0.32	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.05	0.00	0.00	0.00	0.00
ICA with Normal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.87	0.00	0.00	0.00	0.00
ICA with Pearson Type IV	0.00	0.00	0.00	0.00	0.06	0.50	0.10	0.30	0.50	1.00	1.00	1.00	0.00	0.00	0.00	0.01
ICA with <i>t</i> -Location Scale	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22
PCA with Normal	0.00	0.00	0.99	0.52	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.90	0.00	0.00	0.00	0.00
PCA with Pearson Type IV	0.00	0.06	0.48	0.01	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.99	0.00	0.00	0.00	0.04
PCA with <i>t</i> -Location Scale	0.00	0.99	0.11	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.69	1.00	0.00	0.00	0.00	0.00

Source: Euronext. The intraday data come from French Stock Market from January 2002 to December 2006. This table shows the p-value for *DQ* statistics for the alternative VaR models under the 5%, 2.5%, 1%, 0.5% significance thresholds. The *DQ* statistics are asymptotically distributed $\chi^2(7)$; the cells in bold font indicate rejection of null hypothesis of correct VaR estimates at the 5% significance threshold.

Chapter 3

A Robust Conditional Multi-moment Asset Pricing Model with High-frequency Data

3.1 Introduction

The Capital Asset Pricing Model (CAPM) by William Sharpe (1964) and John Lintner (1965) marks the birth of the asset pricing theory; it offers powerful and intuitive predictions about how to measure risk and the relation between expected return and risk. Unfortunately, the empirical record of the model is rather poor. Poor enough to invalidate the way it is used in applications. The CAPM's empirical problems may reflect theoretical failings, as the result of many simplifying assumptions. But they may also be caused by difficulties in implementing valid tests of the model. In this context, due to recent market database availability, several recent research focus on high-frequency data characteristics (see Voit, 2003) and present applications of traditional low-frequency models on newly available high-frequency databases (see Bollerslev and Zhang, 2003), using robust estimation methodology (see Berkowitz and Diebold, 1998).

Indeed, financial variables exhibit strong peculiarities from leptokurticity and asymmetry, to heteroskedascity and clustering phenomenons. Most of classical financial low-frequency models are based on the close-to-normal hypothesis, that is difficult to sustain when real market conditions are under studies. That is the reason why high-frequency data financial applications deserve special research attention and precaution.

In one hand, most of authors now use information contained in the high-frequency series, because being simply closer to the real process is a valuable information (see Kunitomo, 1992), for building denoised lower-frequency estimates of the pertinent parameters that enter into the representative utility function. In the other hand, high-frequency data obviously contain pure noise that has a negative effect of the accuracy of estimations of financial model parameters (see Oomen, 2002). Whilst robust estimators of first and second moments have already been proposed in the literature (see Berkowitz and Diebold, 1998), generalizations in a four-moment world do not yet exist to our knowledge. Similarly, some attention has been recently paid to the conditional modeling of the asset dependences (see Jondeau and Rockinger, 2003) in a heterogeneous market (see Brock and Hommes, 1998 and Malevergne and Sornette, 2006). Based on these ideas, our aim on this chapter is to present an asset pricing model, encompassing some of the most important characteristics of high-frequency financial returns.

We present hereafter some estimations of “realized” measures and co-measures of risk, in the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and in the Arbitrage Pricing Theory (APT) model by Ross (1976) using French stock high-frequency data. Bollerslev and Zhang (2003) demonstrate with US equity transaction data, that the “realized” measures and co-measures are more effective measures of the systematic risk(s) in factor models. Contrary to these standard approaches, we include higher-moments asset pricing models (see Jurczenko and Maillet, 2006), and conditional asset pricing models (see Bollerslev *et al.*, 1988 and Jondeau and Rockinger, 2004).

The motivation for the conditional multi-moment asset pricing with heterogeneous market participants comes from three sources. First, from a theoretical perspective, financial economic considerations suggest that betas may vary with conditioning variables, an idea developed theoretically and empirically in a vast literature (that includes, among many others Berkowitz and Diebold, 1998; Andersen *et al.*, 2002 and Bollerslev and Zhang, 2003). Second, from a different and empirical perspective, the financial econometric volatility literature (see Andersen *et al.*, 2001; Barndorff-Nielsen and Shephard, 2002a and Corsi, 2006) provides extensive evidence of wide fluctuations and high persistence in asset market conditional variances, and in individual equity conditional covariances with the market. Thus, even from a purely statistical viewpoint, market betas, which are ratios of time-varying conditional covariances and variances, might be expected to display persistent fluctuations. Third, all the previous contributions only assume a mean-variance strategy; the investor allocates his portfolio among some risky assets and the risk-free asset. The mean-variance criterion implicitly assumes that returns are normal, or at least that higher moments (beyond mean and variance) are not relevant for the asset allocation.

The outline of the chapter is as follows. Section 3.2 starts with a brief discussion of a general factor pricing model and the notion of realized factor loadings. This section also presents standard summary statistics for the monthly realized portfolio returns and factor loadings for the 43 test portfolios over the 5-year (2002-2006) sample period. Section 3.3 shows our conditional multi-moment asset pricing model with heterogeneous market participants in the high-frequency context. We also propose a robust estimation procedure for this model. Section 3.4 details the case of time-varying investment opportunities, examines the consequences of using the proposal model and provides several robustness checks of our main results. Section 3.5 concludes (with some suggestions for future research).

3.2 Factor Pricing Models and Realized Co-variations

In this section, we introduce the econometric formulations which are considered for the Realized CAPM. The basic idea is that under suitable assumptions about the underlying return generating process, the corresponding factor loading(s) may in theory be estimated arbitrarily well through the use of sufficiently finely sampled high-frequency data (Bollerslev and Zhang, 2003). We describe the return generating process (3.2.1). We then derive the corresponding Realized Loadings (3.2.2), and finally, we discuss the setup of the actual empirical implementation (3.2.3).

3.2.1 Factor Pricing Models

Factor models are amongst the most widely used return generating processes in financial econometrics. They explain co-movements in asset returns as arising from the common effect of a (small) number of underlying variables, called factors. Following common practice in the empirical asset pricing literature, we assume that the underlying discrete-time return generating process is the K -factor model. Specifically, let us denote by \mathbf{R} the N portfolios returns, and by \mathbf{F} the K factors returns. If r_f is the risk free asset return, excess returns of portfolios and factors are defined respectively as: $\mathbf{r} = \mathbf{R} - r_f$ and $\mathbf{f} = \mathbf{F}_t - r_f$. Let \mathbf{r}_i be the excess return of the i -th asset class of the entire portfolio during a specific time interval. The factor model is then specified by:

$$r_i = \alpha_i + \sum_{k=1}^K \beta_{ik} f_k + \varepsilon_i,$$

more compactly:

$$\mathbf{r} = \alpha + \mathbf{B}\mathbf{f}' + \mathbf{e} \quad (3.1)$$

where $\mathbf{B} = (\beta_1, \beta_2, \dots, \beta_K)$ is the $(N \times K)$ loadings, $\mathbf{f} = (f_1, f_2, \dots, f_K)$ is the $(T \times K)$ vector of risk factors, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ is the $(N \times 1)$ vector of intercepts and

$\mathbf{e} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)$ is the $(N \times 1)$ random error with mean 0 accounts for the information not captured by the risk factors. Following equation (3.1), the mean of \mathbf{r} is $E(\mathbf{r}) = \alpha + \mathbf{B}[E(\mathbf{f})]'$ and the variance of \mathbf{r} is $Var(\mathbf{r}) = \mathbf{B}'\Sigma_{\mathbf{F}}\mathbf{B} + \sigma_{\mathbf{e}}^2$ where $\Sigma_{\mathbf{f}}$ and $\sigma_{\mathbf{e}}^2$ are the covariance matrices of \mathbf{f} and \mathbf{e} , respectively. Note that the components β_{ik} of the coefficient vector \mathbf{B} will be zero if the i -th asset class is not exposed to the k -th risk factor.

The factor model is an extension of the well-known market model (Sharpe, 1964; Lintner, 1965), which corresponds to one-factor model (market factor, $r_{M,t}$) with constant loading:

$$\mathbf{r} = \alpha + \beta r_M + \mathbf{e} \quad (3.2)$$

where β is a $(N \times 1)$ vector of market loadings and r_M is a vector of market portfolio returns.

The motivation for using the risks factors of the market is to including most possible information in the model. We can find several sources of bias which are not taken into account in one-factor model. In particular, the liquidity risk and the operational risks which can come from the heterogeneity of information available on the markets. Since there is almost no strictly independent random variable in our economic world, it is almost certain that any additional variable will, at least locally, add to the explanatory power of the CAPM. This form of factor model is most general and all other model can be derived. The factor loadings are then formally defined (see, Bollerslev and Zhang, 2003) by:

$$\mathbf{B} \equiv \frac{cov(\mathbf{r}, \mathbf{f})}{cov(\mathbf{f}, \mathbf{f})} \quad (3.3)$$

This form of estimation of factor betas is generally used for the factor model. But the empirical tests conducted with data from financial markets do not generally imply

the acceptance of the model as describing correctly the range of expected returns. When regressing returns of individual assets on the market factors, R^2 are generally quite low. The regression of the average returns against betas is even worse. Bollerslev and Zhang (2003) propose the use of the realized loadings for estimation of factor loadings.

3.2.2 Realized Loadings and Co-variations

Recently, for the measure of the asset volatility, the use of high frequency data have been advocated to improve the precision of the estimation: the so-called Realized Volatility (RV) approach proposed in a series of breakthrough papers by Andersen *et al.* (2002) and Barndorff-Nielsen and Shephard (2002a and 2002b). As for the Realized Volatility approach, the using high frequency data in the computation of covariances and correlations between two assets leads to the analogous concept of Realized Covariance (or Covariation), Realized Correlation and Realized Betas (Bollerslev and Zhang, 2003).

Following Barndorff-Nielsen and Shephard (2002a), Bollerslev and Zhang (2003) and Morano (2007), we suppose that finest sampled h -period returns are available. Let p_{t-1+ih} denote the $(N \times 1)$ vector of log transaction prices and let $r_{t-1+ih, h} \equiv p_{t-1+ih} - p_{t-1+(i-1)h}$ denote the $(N \times 1)$ vector of returns for the i -th intra-day period on day t , for $i = [2, 3, \dots, 1/h]$, where N is the number of stocks. We assume that $1/h$ is an integer. Then, the corresponding intra-period returns share to the same factor structure as in equation (3.1) with constant intra-period drift and factor loadings such as:

$$r_{t-1+ih, h} = \alpha_t + \beta'_t f_{t-1+ih, h} + \varepsilon_{t-1+ih, h} \quad (3.4)$$

where $f_{t-1+ih, h}$ denote the $(1 \times K)$ vector of returns of the factors for the i -th intra-day period on day t and $\varepsilon_{t-1+ih, h}$ is the random noise vector with $E(\varepsilon_{t-1+ih, h}) = 0$. We have $r_t = \sum_{i=2}^{1/h} r_{t-1+ih, h}$ and $\varepsilon_t \equiv \sum_{i=2}^{1/h} \varepsilon_{t-1+ih, h}$.

The temporally aggregated one-period returns defined by equation (3.4), r_t , is obvi-

ously identically equal to r defined by equation (3.1). With this intra-day period decomposition, Andersen et al. (2002), Barndorff-Nielsen and Shephard (2002a) and Bollerslev and Zhang (2003) suggest to estimate the daily covariation by taking the outer-product of the observed high-frequency returns over the period, namely the Realized Covariance - of course, the true covariance matrix is not directly observable. However, by the theory of quadratic variations¹, the corresponding Realized Covariance (denoted RC) is defined by:

$$RC(r_{t,h}, f_{t,h}) = \sum_{i=2, \dots, 1/h} f_{t-1+ih, h} r'_{t-1+ih, h} \quad (3.5)$$

Andersen et al. (2002) and Protter (2004) show that this statistic converges uniformly in probability to the true covariance matrix as $h \rightarrow 0$. Therefore,

$$RC(f_{t,h}, f_{t,h}) = \sum_{i=2, \dots, 1/h} f_{t-1+ih, h} f'_{t-1+ih, h}, \quad (3.6)$$

is the Realized Volatility of the factors on the day t .

Finally, the conditional betas, or factor loadings, defined in equations (3.1) and (3.4) can be estimated by:

$$\hat{\beta}_{t,h} = \frac{RC(r_{t,h}, f_{t,h})}{RC(f_{t,h}, f_{t,h})} \quad (3.7)$$

$\hat{\beta}_{t,h}$ is the so-called Realized Betas or Factors Loadings (Bollerslev and Zhang, 2003) for the period t using intra-daily sample h . The next subsection presents some statistics of the Realized Betas of stocks returns data on the French Stock Market.

¹See Protter, 2004, for a general discussion

3.2.3 Realized Betas on the French Stock Market

Following the definition of the realized loadings proposed in the previous subsection, we compute here the empirical counterpart on real stocks. We firstly present the characteristics of database and secondly, we present and discuss the factor loadings of some stocks in the French Stock Market.

The data set was obtained from Euronext and consists in transaction prices at the five-minutes sampling frequency for 50 more liquids stocks, covering the period from January 02, 2002 until December 29, 2006 (1,281 trading days). We remove stocks for which the price series start at a later date, leaving 43 stocks for the analysis².

Table 3.1 gives a statistics summary of the selected stocks. These statistics cover minimum and maximum value, mean, standard deviation, skewness, kurtosis and realized volatility (mean and standard deviation of daily realized volatility). We can notice on this table that instantaneous volatilities over all the period are very weak and homogeneous for all entire assets. In the same, the average of the returns is close to zeros. We also see very asymmetrical distributions toward the left (see for example “Pernold-Ricard”, “Sodexho Alliance” and “Air Liquide”) and seldom toward the right-hand side (“Pinault Printemps” and “Vallourec”). Moreover, all the assets have very fat tails distribution (the average kurtosis is higher than 20). These reports show well that it is not possible to support the assumption of normality of the distributions of the returns in the high-frequency context.

The intra-day returns have then been employed in the computation of the daily realized regressions, as discussed in the methodological section, leading to a total of 1,281 daily observations. In order to control the different ranges of variation of the series, standardized variables have been used in realized regression estimations. For obvious reasons, some outliers were also removed from the intra-daily return series.

²See appendix 2 for a list of the company names used.

Table 3.1. Summary Statistics for IntraDaily Returns

	Min	Max	σ	Skew.	Kurt.	Neg. %	Realized Volatility	
							Mean	STD
TF1	-7.3%	8.8%	0.5%	0.54	25.35	47%	7%	5%
EADS	-19.1%	8.5%	0.6%	-1.15	56.92	46%	8%	5%
AIR LIQUIDE	-9.9%	6.1%	0.4%	-1.33	46.54	44%	6%	4%
CARREFOUR	-9.6%	6.8%	0.5%	-0.22	25.42	47%	6%	4%
TOTAL FINA ELF	-6.1%	5.4%	0.4%	-0.22	28.05	44%	5%	3%
OREAL	-5.8%	7.7%	0.4%	0.35	22.97	44%	6%	3%
VALLOUREC	-12.1%	18.5%	0.6%	1.39	63.34	38%	8%	5%
ACCOR	-5.2%	5.0%	0.5%	-0.03	16.90	46%	6%	4%
BOUYGUES	-14.3%	6.8%	0.5%	-0.87	53.41	46%	6%	4%
SUEZ	-11.5%	10.1%	0.6%	0.28	32.22	46%	8%	6%
LAFARGE	-7.7%	6.8%	0.5%	-0.18	24.97	44%	6%	4%
SANOFI SYNTHELABO	-7.8%	6.9%	0.5%	-0.51	26.85	44%	6%	4%
AXA	-8.6%	9.7%	0.7%	0.10	26.73	47%	9%	7%
GROUPE DANONE	-7.7%	7.3%	0.3%	0.93	46.99	43%	5%	3%
PERNOD-RICARD	-20.5%	5.9%	0.4%	-6.25	368.99	42%	5%	3%
LVMH MOET VUITTON	-4.7%	6.4%	0.5%	0.03	16.50	43%	6%	4%
SODEXHO ALLIANCE	-22.2%	8.7%	0.6%	-2.74	131.81	47%	7%	5%
MICHELIN	-5.4%	4.6%	0.5%	0.11	12.85	46%	6%	3%
THALES	-5.3%	6.2%	0.5%	0.24	15.25	47%	6%	4%
PINAULT PRINTEMPS	-6.7%	17.4%	0.5%	2.02	85.11	44%	6%	5%
PEUGEOT	-7.5%	5.7%	0.5%	-0.41	24.42	47%	6%	4%
ESSILOR INTL	-4.4%	7.3%	0.4%	0.46	18.11	44%	6%	3%
SCHNEIDER ELECTRIC	-5.5%	4.5%	0.4%	-0.08	16.28	44%	6%	4%
VIVENDI ENVIRON.	-7.2%	6.3%	0.5%	0.09	23.49	47%	7%	5%
SAINT-GOBAIN	-12.9%	7.5%	0.5%	-0.70	36.80	46%	7%	5%
CAP GEMINI	-16.2%	15.8%	0.7%	0.09	42.56	48%	9%	7%
CANAL +	-6.4%	5.9%	0.5%	-0.15	15.71	32%	7%	5%
VINCI	-3.3%	5.2%	0.4%	0.46	14.13	42%	5%	3%
CASINO GUICHARD	-6.6%	6.2%	0.4%	-0.04	21.24	43%	6%	3%
AGF	-7.3%	11.8%	0.5%	0.57	31.71	44%	7%	5%
VIVENDI UNIVERSAL	-18.3%	13.7%	0.8%	-1.34	67.74	47%	9%	8%
DEXIA SICO.	-9.0%	7.6%	0.5%	-0.05	37.45	43%	6%	5%
STMICROELEC.SICO.	-7.4%	7.4%	0.6%	0.18	22.91	46%	8%	5%
ALCATEL A	-21.1%	15.4%	0.9%	-0.57	52.82	45%	11%	9%
LAGARDERE	-4.1%	5.7%	0.5%	0.33	14.34	44%	6%	4%
VALEO	-4.6%	7.7%	0.5%	0.63	18.97	47%	7%	4%
PUBLICIS GROUPE	-6.8%	8.4%	0.5%	0.30	20.36	47%	7%	5%
SOCIETE GENERALE A	-7.4%	8.3%	0.5%	0.34	26.68	43%	7%	5%
BNP PARIBAS	-8.0%	8.0%	0.5%	0.35	27.00	45%	7%	4%
RENAULT	-6.1%	8.5%	0.5%	0.27	22.64	44%	7%	4%
FRANCE TELECOM	-16.3%	15.7%	0.8%	-0.79	66.70	47%	9%	8%
THOMSON	-10.8%	11.1%	0.6%	0.26	31.08	46%	8%	5%
CREDIT AGRICOLE	-15.0%	7.3%	0.5%	-1.36	63.98	46%	6%	4%

Source: Euronext. The table reports some main statistics of the intra-daily returns. This statistics cover minimum (Min) and maximum value (Max), standard deviation (σ), skewness (Skew.), kurtosis (Kurt.), negative returns frequency (Neg.) and realized volatility (mean and standard deviation of daily realized volatility) over the January 2002 through December 2006 sample period. The realized daily volatilities are reported in standard deviation format, and constructed from the summation of the squared 5-min returns. The mean of all stocks 5-min returns is close to 0.

For the empirical estimation, the 5-min compounded returns on the market portfolio (denoted MKT), are constructed as the logarithmic transform of the equally-weighted percentage returns across all of the stocks. Following the approach in Fama and French (1993), the excess returns for the style portfolio (denoted HML) and the size portfolio (denoted SMB) factor are defined by the return differential between the sorted portfolios grouped according to the individual stocks book-to-market ratios and their market capitalization respectively. We also compute 9 portfolios by merging 3 portfolios according to the size with 3 portfolios according to the style (for instance the portfolio “11” will be computed by the average of the asset returns being in the 1st third sorted by the market capitalization and 1st third sorted by book-to-market). In Table 3.2, summary statistics of the monthly returns for the 9 merged portfolios are presented. We observe that all portfolios has positively average of monthly returns. Also value growth stocks (low book-to-market value) in the small size outperformed growth stocks (high book-to-market value) of the same size. Meanwhile, the average realized volatilities over the recent 5-year period are generally in line with the long-run historical sample standard deviations, with the portfolios in the lowest book-to-market exhibiting the highest average return volatility.

Table 3.2. Mean, Realized Volatility for Monthly Returns

	Low	Middle	High
Mean Returns			
Small	0.055	0.068	0.093
Middle	0.061	0.050	0.084
Big	0.064	0.060	0.079
Realized Volatility			
Small	0.43	0.14	0.41
Middle	0.37	0.18	0.36
Big	0.78	0.13	0.38

Source: Euronext. The table reports the average monthly percentage returns, realized volatility, average realized factor loadings for the 9 size and book-to-market sorted portfolios over the January 2002 through December 2006 sample period. The realized monthly volatilities are reported in standard deviation format, and constructed from the summation of the squared 5-min returns.

Although the most liquid stocks in our sample generally is traded several times within each 5-min interval, some of the less liquid stocks may not be traded for up to an hour or longer. As such, the 5-min portfolio returns are clearly under the influence of non synchronous trading effects. It is well known that such effects may systematically bias the estimates of the factor loadings from traditional time series regressions. Several different adjustment procedures have been proposed in the empirical asset pricing literature for dealing with this phenomenon over coarser daily and monthly frequencies (see, e.g., Campbell *et al.*, 1997). In particular, under the simplifying assumption that the trading intensity and the true latent process for the returns are independent, the one-factor CAPM adjustment procedure first proposed by Scholes and Williams (1977) works by accumulating additional leads and lags of the sample auto-covariances between the returns and the market portfolio. Adapting the Scholes-Williams procedure to a multi-factor framework, our empirical variation and covariation measures for the factors and the returns are based on the following generalizations of equations (3.5) and (3.6):

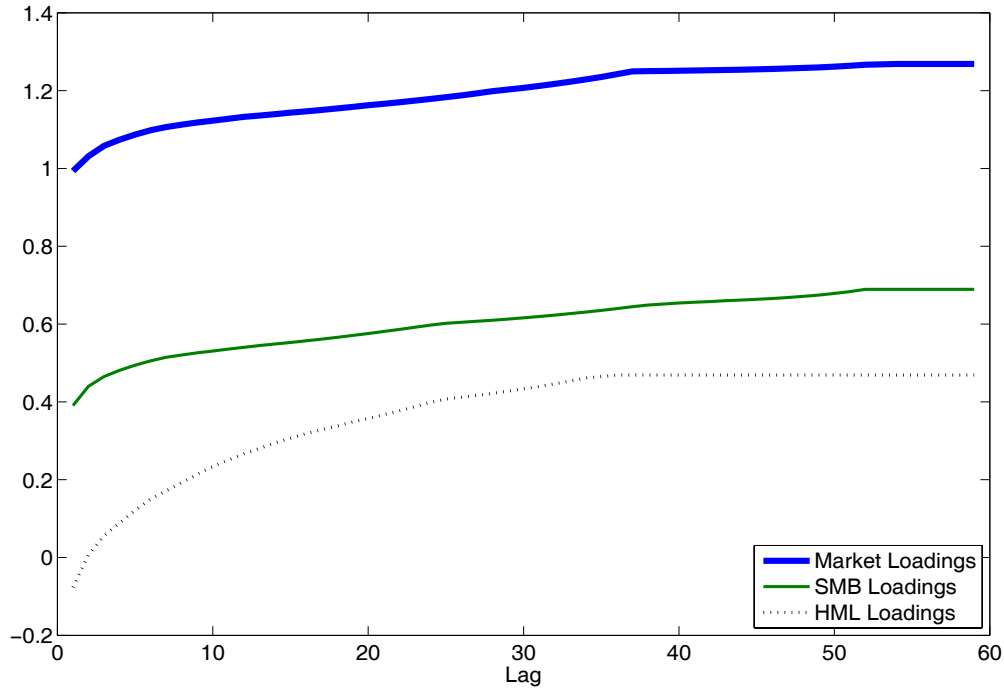
$$RC_L(r_{t,h}, f_{t,h}) = \sum_{j=-L}^L \left[\sum_{i=1}^{1/h} f_{t+ih,h} f'_{t+(i-j)h,h} - 2Lh \left(\sum_{i=1}^{1/h} f_{t+ih,h} \right) \times \left(\sum_{i=1}^{1/h} r_{t+(i-j)h,h} \right)' \right] \quad (3.8)$$

where L denotes the maximum lag length included in the adjustment. The adjusted realized factor loadings are then simply obtained by direct substitution of $RC_L(r_{t,h}, f_{t,h})$ and $RC_L(f_{t,h}, f_{t,h})$ in place of $RC(r_{t,h}, f_{t,h})$ and $RC(f_{t,h}, f_{t,h})$ respectively in equation (3.7)³.

We choose a lag window of $L = 6$ in the empirical results reported on below. This particular choice of L was motivated by visual inspection of the adjusted factor loadings, which showed little sensitivity to the choice of L beyond that lag. To illustrate this feature, Figure (3.1) plots the average factor loadings across the full sample associated with the Market, Small Minus Big (SMB) size and High Minus Low (HML) style factors

³Note that for $L = 0$ the expressions reduce to the standard expressions underlying the conventional unadjusted OLS regression estimates.

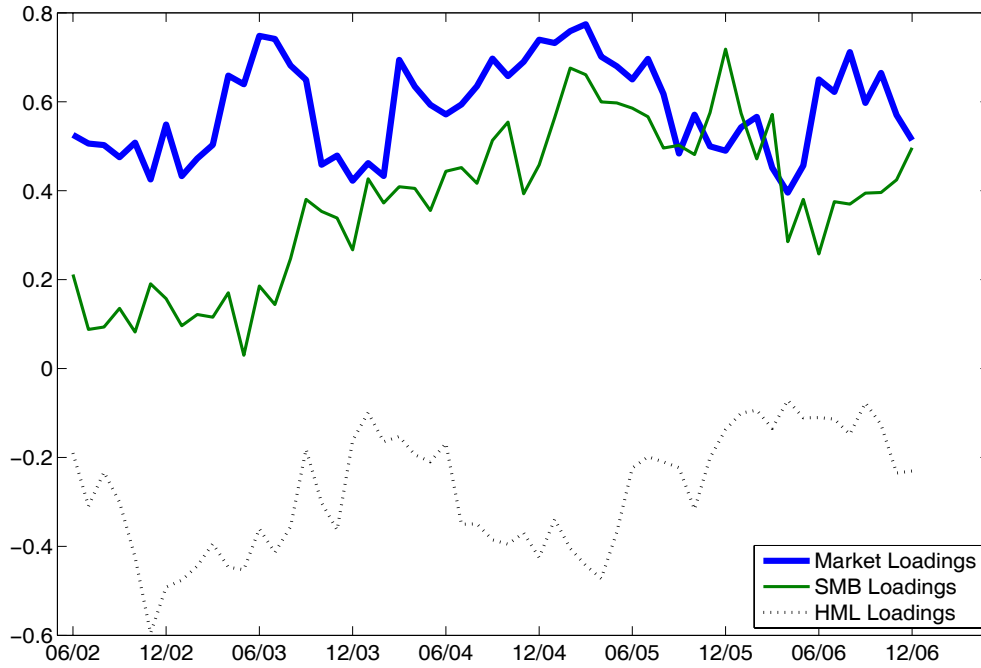
Figure 3.1. Average Adjusted Factor Loadings for smaller size and style Portfolio



Source: Euronext. The database includes a complete set of five-minutes transactions prices from January 02, 2002, through December 29, 2006 (1,281 trading days). The figure plots the average Scholes-Williams adjusted factor loadings for smaller size and style Portfolio as a function of the adjustment, $L = [0, 1, 2, \dots, 10]$. SMB corresponds to the Small Minus Big size factor and HML is the High Minus Low style factor.

representing portfolios for the test as a function of the lag length, L and Figure 3.2 plots the factors loadings with lag 6. The smaller size and style Portfolio consists of the stocks in the first size and first book-to-market value, and as such is among the test portfolios most prone to non-synchronous trading effects. Nonetheless, it is evident from the figure that the average loadings for all three factors stabilize fairly quickly, and is very close to the average full-day lag 59 adjusted loadings for L around 6. A similar picture emerges for all of the other portfolios with the convergence to the full-day average occurring even faster (smaller values of L) in most other cases.

Figure 3.2. Realized Adjusted Factor Loadings of Large Size and Higher Style Portfolio



Source: Euronext. The database includes a complete set of five-minutes transactions prices from January 02, 2002, through December 29, 2006 (1,281 trading days). The figure plots the average Scholes-Williams adjusted factor loadings with lag 6 for large size and higher style Portfolio. SMB corresponds to the Small Minus Big size factor and HML is the High Minus Low style factor.

3.3 Conditional Asset Pricing Model

In this section we introduce the econometric formulations that are considered in this chapter. In turn, we describe the return generating process. We derive the regressions tests of the CAPM Model, and finally we propose our framework estimation.

3.3.1 Realized Conditional Asset Pricing Model with Higher Moments

We include in the return generating process the third moment and also the intra-daily variabilities. We first present the conditional asset pricing with three moments and after,

we show the corresponding realized measures.

Three-moment Conditional Asset Pricing

Recall \mathbf{r} the $(N \times T)$ matrix of excess returns of N stocks; $\mathbf{r} = (r_1, r_2, \dots, r_N)$ and \mathbf{f} the $(K \times T)$ excess return of risky factors; $\mathbf{f} = (f_1, f_2, \dots, f_K)$. We have:

$$r_i = \alpha_i + \sum_{k=1}^K \beta_{ik} f_k + \sum_{k=1}^K \gamma_{ik} f_k^* + \varepsilon_i, \quad (3.9)$$

where $f_k^* = F_k^2 - r_f$, is the excess returns of the square returns of factor k , with $k = [1, \dots, K]$; f_k is the k -th common factor, β_{ik} is the loadings of stock i on each of the k factors, γ_{ik} is the sensitivity of stock i on each of the k quadratic factors f_k^* and ε_i are assumed to be uncorrelated across assets.

More compactly:

$$\mathbf{r} = \alpha + \mathbf{B}\mathbf{f} + \mathbf{G}\mathbf{f}^{*'} + \mathbf{e} \quad (3.10)$$

where \mathbf{B} is $(N \times K)$ matrix of loadings, \mathbf{G} $(N \times K)$ matrix of loadings of the quadratic factors and \mathbf{e} is the $(T \times N)$ matrix of residuals with zero mean and covariance matrix Σ .

At equilibrium, the expected return of an asset i is linearly associated with the contributions of the asset to the variance and skewness of the risk factor. The conditional version of this multi-factor model (Krauss and Litzenberger, 1976; Barone *et al.*, 2002; Smith, 2007) is:

$$E(r_i) = \lambda_0 + \sum_{k=1}^K (\beta_{ik} \lambda_{1,k} + \gamma_{ik} \lambda_{2,k}) \quad (3.11)$$

where λ_0 may be interpreted as the expected excess returns of a portfolio with zero covariance and coskewness with the market, $\lambda_{1,k}$ is the price of beta risk, $\lambda_{2,k}$ is the price of gamma risk. The value of the loading factor is given by equation 3.3 and:

$$\gamma_{i,k} = \frac{\text{Coskew}(r_i, f_k)}{\kappa_j^3}, \quad (3.12)$$

is the conditional gamma of asset i . The conditional coskewness of the asset i with the factor k , $\text{Coskew}(r_i, f_k)$ is given by:

$$\text{Coskew}(r_i, f_k) = E[(r_i - \bar{r}_i)(f_k - \bar{f}_k)^2], \quad (3.13)$$

where $\bar{r}_i = E(r_i)$ and $\bar{f}_k = E(f_k)$ and the conditional coskewness of the factor k , κ_k^3 is as follows:

$$\kappa_k^3 = E[(f_k - \bar{f}_k)^3] \quad (3.14)$$

The motivation for including the square of the market returns is to fully account for coskewness with the factors. The quadratic market model (equation 3.10) is a direct extension of the well-known market model (Sharpe, 1964 and Lintner, 1965), which corresponds to some restrictions. With this return generating process, we assume that the individual error terms are correlated; so we have non-zero covariances between the error terms ε_i . This special type of problems is called in the econometric literature the Seemingly Unrelated Regressions (SUR) Model (Zellner, 1962). We propose some basic intuition of the estimation procedure and some extensions (Parks, 1967) in the special subsection of this section.

Realized Loadings in Conditional Asset Pricing with Three moments

We develop and extend in this subsection the conditional multi-factor Asset Pricing Model by adding unobserved prices variabilities in the low frequency. Suppose as in the previous section that finer sampled h -period returns are available. The h -period three-moment conditional asset pricing is therefore given, in compact form similar to equation 3.4:

$$r_{t-1+ih, h} = \mu_t + \beta'_t f_{t-1+ih, h} + \gamma'_t f_{t-1+ih, h}^* + \epsilon_{t-1+ih, h} \quad (3.15)$$

where a new parameter γ_t is $(K \times 1)$ vector of conditional gamma at the time t . The conditional betas are defined by equation 3.7.

The Realized Coskewness (denoted RCoskew) of assets and risks factors with h -period

sampling is defined (Beine *et al.*, 2004) by:

$$\text{RCoskew}(r_{t,h}, f_{t,h}) = \sum_{i=2}^{1/h} r_{t,h} f_{t,h}^* = \sum_{i=2}^{1/h} r_{t,h} (F_{t,h}^2 - R_{Ft,h}) \quad (3.16)$$

Therefore, the Realized Gamma ($\hat{\gamma}_{t,h}$) for the day t is:

$$\hat{\gamma}_{t,h} = \frac{\text{RCoskew}(r_{t,h}, f_{t,h})}{\text{RCoskew}(f_{t,h}, f_{t,h})} \quad (3.17)$$

3.3.2 Regressions Tests of the CAPM

We describe here two types of the most popular regression tests of CAPM. The first group of tests of the Capital Asset Pricing Model is related to time-series variation (Black *et al.*, 1972 and Gibbons *et al.*, 1989) whilst the second is based on cross-sectional regressions (Fama and Macbeth, 1973).

Times-series Regressions

Recall the factor pricing model with a single factor⁴. All variables hereafter are considered in terms of excess returns. The loading for the asset i is given by the regression:

$$r_i = \alpha_i + \beta_i f + \varepsilon_i \quad (3.18)$$

The model states that expected returns are linear in the betas: $\bar{r}_i = \beta_i \bar{f}$. Then this means that all the regression intercepts α_i should be zero. Black *et al.* (1972) and Gibbons *et al.* (1989) developed a methodology for this test. That we briefly present in the next subsection.

The Black, Jensen and Scholes (BJS) Method

Black *et al.* (1972) suggested a natural strategy for estimation and evaluation. Run time-

⁴For simplify notations, we use here a single factor model; but the results are the same for multi-factor case with vectors in place of scalars.

series regressions (equation 3.18) for each test. The estimate of the factor risk premium is just the sample mean of the factor: $\hat{\lambda} = \bar{f}$. Then, for the case that the regression errors in (equation 3.18) are uncorrelated and homoskedastic, use standard Ordinary Least Square (OLS) *formula* for a distribution of the parameters can be used⁵. We can also test the case that all the pricing errors are jointly equal to zero. With errors that are independent and identically distributed over time, homoskedastic and independent of the factors, the asymptotic joint distribution of the intercepts gives the model test statistic (χ^2 test, see Campbell *et al.*, 1997 and Cochrane, 2001),

$$t_N = T [1 + (\bar{f}\hat{\sigma}_f^{-1})^2]^{-1} \hat{\alpha}\hat{\Sigma}^{-1}\hat{\alpha} \quad (3.19)$$

where \bar{f} denotes sample mean of factor f , $\hat{\sigma}_f^2$ denotes sample variance of factor f , $\hat{\alpha}$ is a vector of the estimated intercepts, $\hat{\Sigma}$ is the residual covariance matrix: $\hat{\Sigma} = E(\hat{\mathbf{e}}\hat{\mathbf{e}}')$. t_N converges asymptotically to χ_N^2 .

For the BJS Wald-type test defined by equation (3.19), since Σ is unknown, its approximation needs the large-sample distribution theory for drawing inference (Campbell *et al.*, 1997 and Cochrane, 2001). Gibbons *et al.* (1989) give a finite-sample distribution for the test.

The Gibbons, Ross and Shanken (GRS) Method

The candidate model for times-series regressions of Gibbons *et al.* (1989) is the Market Model regression equation (3.18). Similar to univariate F -test (which is equal to a squared t -test), its multivariate analogue looks at the ratio between the square of the estimated coefficients and their variance; in matrix notation, this involves the quadratic expression:

$$\hat{\alpha}'[Var(\hat{\alpha})]^{-1}\hat{\alpha} = T(\hat{\alpha}'\Sigma^{-1}\hat{\alpha}) [1 + (\bar{f}\sigma_f^{-1})^2] \quad (3.20)$$

⁵The t -tests is use to check whether the pricing errors α are in fact zero.

where Σ is the variance-covariance matrix of the residuals in equation (3.18). The finite-sample test of Gibbons *et al.* (1989) is achieved by an appropriate adjustment of the above expression for the degree of freedom. The test statistic W is defined as:

$$W = (T - N - 1)N^{-1} [1 + (\bar{f}\sigma_f^{-1})^2]^{-1} \hat{\alpha}'\Sigma^{-1}\hat{\alpha} \quad (3.21)$$

They show that, under the null hypothesis, W is unconditionally distributed as a central F with N degrees of freedom in the numerator and $(T - N - 1)$ degrees of freedom in the denominator⁶, that is:

$$W \sim F_{N, T-N-1} \quad (3.22)$$

The GRS test statistic as a useful economic interpretation since Gibbons *et al.* (1989) show that this statistic can be rewritten in terms of Sharpe ratios:

$$W = (T - N - 1)N^{-1} \left[\frac{E(q)^2}{\sigma_q^2} - \frac{\bar{f}^2}{\sigma_f^2} \right] \left\{ 1 + [\bar{f}\sigma_f^{-1}]^2 \right\} \quad (3.23)$$

where the portfolio denoted by q is the ex-post tangency portfolio constructed from the N assets plus the test factor f . The GRS test statistic W can therefore be interpreted in terms of the difference in squared Sharpe ratios between q and f . If f is *ex post* efficient, it has the highest Sharpe ratio of all portfolios which can be constructed and it will have the same Sharpe ratio as q . Otherwise, q can only have a higher Sharpe ratio⁷ (Campbell *et al.*, 1997; Cochrane, 2001).

If there are many factors that are excess returns, the quadratic form $(\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha})$ has the distribution:

$$\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} = (T - N - K)N^{-1} \left[1 + \bar{f}'\hat{\Omega}^{-1}\bar{f} \right]^{-1} \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} \sim F_{N, T-N-K} \quad (3.24)$$

where N is the number of asset, K the number of factors and $\hat{\Omega} = \frac{1}{T} \sum_{t=1}^T (f - \bar{f})(f - \bar{f})'$

⁶See Campbell *et al.* (1997) and Cochrane (2001) for more details.

⁷We suppose here, for our analysis of the economic perspective of GRS test statistic, that the factor f is the Market portfolio.

is the sample covariance matrix of factors.

Cross-sectional Regressions (Fama and Macbeth, 1973)

The CAPM imposes a cross-sectional restriction, namely that expected asset returns are linearly related to their market beta; it thus seems to be straightforward to test the CAPM by running a cross-sectional regression of expected returns on betas. If we want to test specific alternatives, we could just add additional explanatory variables to the regression. In a nutshell, that is already the basic idea of cross-sectional regressions tests. But we do not know expected returns and betas, we can only estimate their values. When running cross-sectional regressions, we also need to account for potential cross-sectional correlation of residual returns. While the assumption of zero serial correlation is generally a good approximation for monthly returns, abnormal returns do considerably co-vary across assets. The cross-sectional regression model specified by Fama and MacBeth (1973) have the form for each period t ⁸:

$$r_{i,t} = \lambda_{0,t} + \lambda_{1,t}\beta_{i,M|t} + \sum_{j=2}^J \lambda_{j,t}k_{i,j|t} + \xi_{i,t} \quad (3.25)$$

over all assets i , with $i = [1, \dots, N]$, and T observed periods with $t = [1, \dots, T]$. Where $\lambda_{0,t}$ is the return in excess of the risk free rate on a fully-invested, least-variance portfolio where the effects of the others variables are zero, $\lambda_{1,t}$ is the return of a zero-investment, least-variance portfolio with a beta of one and $\lambda_{j,t}$ (for $j > 2$) is the return on a zero-investment, least-variance portfolio where the variables k_j are zero⁹.

If the CAPM holds, the expected return of $\lambda_{j,t}$ (for $j \neq 1$) should be zero, the expected return of $\lambda_{1,t}$ should be positive. Using time-series averages as estimates of

⁸The Fama-MacBeth method allows to incorporate time-varying betas.

⁹The relation equation (3.25) insures that there are no other parameters next to beta which explain the cross-section of expected returns. For instance, Fama and MacBeth (1973) test whether expected returns are also determined by variance. They postulate an alternative relationship between expected returns, beta and idiosyncratic risk $\sigma(\varepsilon_{i,t})$ of model defined in equation (3.18)

expected values, we can test whether these are significantly different from zero with standard t -tests¹⁰. We have for $j = [1, \dots, N]$:

$$\begin{cases} \bar{\lambda}_j = T^{-1} \sum_{t=1}^T \hat{\lambda}_{j,t} \\ \hat{\sigma}(\bar{\lambda}_j) = \left\{ [T(T-1)]^{-1} \sum_{t=1}^T (\hat{\lambda}_{j,t} - \bar{\lambda}_j)^2 \right\}^{1/2} \\ \hat{t}(\bar{\lambda}_j) = \bar{\lambda}_j [\hat{\sigma}(\bar{\lambda}_j)]^{-1} \end{cases} \quad (3.26)$$

Cochrane (2001) shows how to embed the Fama-MacBeth method in a more general estimation context.

3.3.3 Estimation Procedure of the Model

For a robust estimation of the proposed model, we combine Seemingly Unrelated Regressions (SUR) Zellner (1962) and cross-sectionally correlated of Parks (1967) methodologies. This because the Seemingly Unrelated Regressions (SUR) model, proposed by Zellner (1962), can be viewed as a special case of the generalized regression model, but it does not share all of the features or problems of other leading special cases (e.g., models of heteroskedasticity or serial correlation). Parks (1967) proposes an estimation procedure that allows the error term to be both serially and cross-sectionally correlated. Consider a system of N simultaneous equations, of which the typical i -th equation defined in equation 3.1. More compactly, those latter vectors can be stacked into an NT -dimensional vector r , with a corresponding arrangement for the error terms, coefficient vectors, and regressors:

$$\mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix},$$

¹⁰The test is formulated as:

$$\begin{cases} \bar{\lambda}_0 = 0 \\ \bar{\lambda}_1 > 0 \\ \bar{\lambda}_j = 0, \text{ for } j > 1 \end{cases}$$

\mathbf{r} , \mathbf{e} and \mathbf{B} are respectively an $(NT \times 1)$ vector of returns, $(NT \times 1)$ vector of errors terms and $(K \times 1)$ vector of loadings, and:

$$\mathbf{F} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_N \end{pmatrix},$$

with $K = \sum_{i=1}^N k_i$ and \mathbf{F} the $(NT \times K)$ matrix of factors. With this notation, we have:

$$E(\mathbf{r}) = \mathbf{B}\mathbf{F} \quad (3.27)$$

Therefore, given that e_{it} is the error for the i -th asset in the t -th time period, the assumption of contemporaneous disturbance correlation, but not correlation over time, implies that the covariance matrix of this SUR system is:

$$V(\mathbf{r}) = \mathbf{\Sigma} \otimes \mathbf{I} \quad (3.28)$$

where:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{pmatrix},$$

with \mathbf{I} the $(T \times T)$ identity matrix and \otimes the *Kronecker product* notation indicating that each element of $\mathbf{\Sigma}$ is multiplied by an identity matrix.

With the previous notations, the classical Least Squares estimator for the vector \mathbf{B} is:

$$\hat{\mathbf{B}}_{\text{LS}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{r} \quad (3.29)$$

and the GLS estimator (assuming Σ is known) is:

$$\hat{\mathbf{B}}_{\text{GLS}} = \left[\mathbf{F}'(\Sigma \otimes I)^{-1} \mathbf{F} \right]^{-1} \mathbf{F}'(\Sigma \otimes I)^{-1} \mathbf{r} \quad (3.30)$$

Now assume that the elements of the disturbance vector follow an $AR(1)$ process (see Parks, 1967 and Bali and Wu, 2005):

$$e_{it} = \rho_i e_{it-1} + \epsilon_{it}, \quad (3.31)$$

where $|\rho_i| < 1$ and ϵ_{it} are random variables with zero mean. Equation (3.31) indicates that the e_{it} are correlated. Equation (3.1) for asset i can then be written as:

$$r_i = \mathbf{b}_i \mathbf{F}_i + \mathbf{P}_i \epsilon_i \quad (3.32)$$

with $e_i = \mathbf{P}_i \epsilon_i$, where ϵ_i is a $T \times 1$ random vector with $E(\epsilon_i) = 0$, $E(\epsilon_i \epsilon_i') = \sigma_{ii} I$ and:

$$\mathbf{P}_i = \begin{bmatrix} (1 - \rho_i^2) - 1/2 & 0 & 0 & \cdots & 0 \\ \rho_i(1 - \rho_i^2) - 1/2 & 1 & 0 & \cdots & 0 \\ \rho_i^2(1 - \rho_i^2) - 1/2 & \rho_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_i^{T-1}(1 - \rho_i^2) - 1/2 & \rho_i^{T-2} & \rho_i^{T-3} & \cdots & 1 \end{bmatrix} \quad (3.33)$$

Under this setup, Parks (1967) presents a consistent and asymptotically efficient three-step estimation technique for the regression coefficients. The first step uses single equation regressions to estimate the parameters of autoregressive model. The second step uses single equation regressions on transformed equations to estimate the contemporaneous covariances. Finally, the Aitken estimator is formed using the estimated covariance,

$$\hat{\mathbf{B}} = (\mathbf{F}' \Omega^{-1} \mathbf{F})^{-1} \mathbf{F}' \Omega^{-1} \mathbf{r} \quad (3.34)$$

where $\Omega = E(\mathbf{e} \mathbf{e}')$ denotes the general covariance matrix of the innovation.

These estimators of $\mathbf{\Omega}$ are not unbiased for true covariance matrix, but they are consistent under the usual conditions. Again imposing reasonable regularity conditions, it can be shown that this \mathbf{B} estimator is asymptotically equivalent to the infeasible Generalized Least Square (GLS) estimator which assumes $\mathbf{\Sigma}$ is known. So inference on the parameter vector \mathbf{B} can be carried out using the approximate normality of $\hat{\mathbf{B}}$:

$$\hat{\mathbf{B}} \sim IN(\mathbf{B}, \mathbf{\Omega}) \quad (3.35)$$

In the application, we use this mentioned methodology with the slope coefficients restricted to be the same for all assets. In particular, we use the same three-step procedure and the same covariance assumptions as in (equation 3.31) to (equation 3.33) to estimate the covariances and to generate the t -statistics for the parameter estimates.

3.4 Results and Robustness Tests

In this section, we examine the pricing efficiency of alternative models. The Fama-French Realized Portfolio Returns and Asset pricing Test are presented. We firstly present the gains associated with the use of high frequency data in the estimation of factor loadings and asset pricing. We then add a realized coskewness in the model and compare with the Bollerslev and Zhang (2003) realized loading procedure.

3.4.1 Fama-French Realized Portfolio Returns and Factor Loadings

The gains associated with the use of the high-frequency data for more accurately measuring and forecasting the realized factor loadings depend upon the quality of the corresponding asset pricing model. The empirical results in the literature discussed above suggest that the asset pricing anomalies related to size and book-to-market largely disappear within the context of the three-factor Fama and French (1993) model. As such, this particular model and the 9 tests portfolios sorted on the basis of firm size and book-

to-market ratios provide a natural setting for evaluating the improvements entailed by the high frequency-based factor loadings¹¹.

The average monthly percentage loadings and standard deviations, over the 5-year (2002-2006) sample period are given in Table 3.3. In contrast to the earlier results, the cross-sectionally averages no longer show a clear-cut pattern in regards to the size and book-to-market characteristics. Indeed, as noted below, the model still yields an average R^2 in excess of 30% for explaining the cross-sectional differences in the monthly returns across the 9 portfolios more than twice the average R^2 for the standard one-factor CAPM over the same time period. Also, the average R^2 for the 9 time series regressions of the portfolio returns on the three factors exceeds 75%. These results are directly in line with the results reported in the existing literature based on longer historical samples.

Table 3.3. Realized Factor Loadings

Mean of Realized Factor Loadings									
	Market			Small Minus Big			High Minus Low		
	Low	Middle	High	Low	Middle	High	Low	Middle	High
Small	1.098	0.708	0.638	0.505	0.418	0.259	0.150	-0.067	-0.223
Middle	1.020	0.696	0.656	0.532	0.424	0.268	0.071	-0.063	-0.253
Big	0.973	0.687	0.576	0.631	0.436	0.389	0.072	-0.057	-0.278
Standard Deviation of Realized Factor Loadings									
Small	0.131	0.035	0.070	0.132	0.057	0.123	0.189	0.057	0.259
Middle	0.066	0.036	0.076	0.074	0.052	0.115	0.134	0.049	0.130
Big	0.096	0.031	0.071	0.082	0.050	0.154	0.148	0.049	0.123

Source: Euronext. The table reports the average realized factor loadings for the three-factor Fama and French (1993) Model for each of the 9 size and book-to-market sorted portfolios over the January 2002 through December 2006 sample period and the corresponding sample standard deviations. The realized factor loadings are constructed from high-frequency 5-min returns.

The average loadings for the Small Minus Big (SMB) size factor systematically decrease with the size of the portfolio, while the average realized loadings for the High

¹¹The 9 portfolios is computed by merging 3 portfolios according to the size with 3 portfolios according to the style (for instance the portfolio “11” will be computed by the average of the asset returns being in the first third sorted by the market capitalization and first third sorted by book-to-market).

Minus Low (HML) factor increase with the book-to-market ratio. No apparent pattern is evident in the loadings association with the return on the Market portfolio (MKT). The monthly sample standard deviations reported suggest that the realized factor loadings also vary a lot through time, with the loadings for the largest low book-to-market portfolios. This will inflate the reported standard deviations. It would be interesting, but beyond the scope of the present chapter, to formally assess the magnitude of the measurement errors in the present context (Bai *et al.*, 2000; Barndorff-Nielsen and Shephard, 2002a and 2002b and Meddahi, 2002; all authors providing important theoretical results for the case of scalar realized volatility measures). Even though the inherent measurement errors result in somewhat erratic month-to-month realized factor loadings, an underlying systematic temporal pattern is evident for all three risk factors. Similar results are true for the other portfolios.

Table 3.4. Average R^2 for Alternatives Models with or without Skewness Factor

	Realized Regression			Classical Regression		
	Low	Middle	High	Low	Middle	High
Without Skewness Factor						
Small	59.33%	85.12%	65.22%	34.69%	84.23%	30.31%
Middle	52.01%	85.94%	64.69%	38.70%	84.59%	33.62%
Big	59.59%	85.86%	75.86%	36.32%	83.99%	28.73%
With Skewness Factor						
Small	77.03%	95.34%	88.85%	47.57%	85.35%	36.07%
Middle	72.21%	96.10%	89.01%	47.44%	86.21%	37.52%
Big	73.65%	95.71%	88.96%	46.15%	86.09%	32.42%

Source: Euronext. The table reports the average of R^2 for the three-factor Fama and French (1993) Model and Fama and French (1993) Model with squared factor returns for each of the 9 size and book-to-market sorted portfolios over the January 2002 through December 2006 sample period. The estimations are carried out on realized regression and on classical static regression.

The squared factor is added to the three-factor Fama and French (1993) Model. The Table 3.4 gives the average of R^2 for the three-factor Fama and French (1993) Model and Fama and French (1993) Model with Squared factor returns for each of the 9 size and book-to-market sorted portfolios over the January 2002 through December 2006 sample

period. The alternatives estimations are also made in realized regression and classical static regression. We observe clearly that the level of the R^2 is high with the squared factor that others estimations methods. This means the addition of the skewness factor makes the classical three-factor model strikingly more competitive. For example, in the large size and high style portfolio, the classical three-factor model produces an R^2 of 28.73 percent, the realized estimation gives 65.86 percent when the model with skewness factor produces 88.96 percent. The intercepts of the regressions for this last model is also relatively weak compared to the other models¹².

The realized factor loadings are, of course, only observable *ex post*. The results in the next section demonstrate that a simple $AR(1)$ time series model for the high-frequency factor loadings effectively smoothes out the measurement error, and gives rise to more accurate forecasts and statistically significant lower pricing errors when compared to the conventional monthly rolling regression-based estimates traditionally used in the literature.

3.4.2 High-frequency Factor Loadings and Asset Pricing

The discussion in the previous section illustrated how high-frequency data may be used in an actual empirical setting for more accurately measuring systematic factor risk exposures. This section goes one step further and asks the question: do the high frequency-based measures give more accurate predictions of the future factor loadings and statistically significant improvements in terms of asset pricing predictions and corresponding better economic decisions? We begin by a discussion of our four different return-based forecast procedures for the factor loadings to based on monthly return observations only.

¹²We also tested the model with the *kurtosis*, but this factor is not significant and was not finally retained for the tests

Predicted Factor Loadings

For the predicted factor loadings, we use several independent procedures. We first considers the factor loadings to be constant and equal to the conventional regression estimates based on 5 years of monthly observations at the beginning of the out-of-sample forecast comparison period. We will refer to these Constant factor loadings by the acronym CON. We have also experimented with the loadings estimates based on the monthly return observations over the full sample. The 5-year Rolling Regression-based (RR) estimation technique advocated by Fama and MacBeth (1973) allows the betas to change every month with the new return observations. As previously noted, on using this procedure along with the actual realized factor returns, the average of the 60 monthly cross-sectional R^2 's (for explaining the variation in the returns across the 9 portfolios) equals 0.929, whereas the average R^2 from the 9 time series regressions equals 0.756.

We use, as a benchmark, a high-frequency-based procedure takes the forecast for the next month to be equal to the current months realized loadings. These forecasts would, of course, be optimal in a Mean Square Error sense if the time series of realized factor loadings followed a Random Walk (RW), or martingale, model. We consequently refer to this procedure as the HF-RW-based forecasts. The second high-frequency procedure is based on the one-step-ahead forecasts from a set of $AR(1)$ models for each of the different loadings and portfolios. The estimates of the models are based on a rolling 24-months sample, leaving us with an out-of-sample forecast comparison period covering the 60 months from January 2002 through December 2006. Consistent with the significant autocorrelations reported in Table 3.5, the vast majority of the parameter estimates for the $3 \times 9 \times 60 = 1620$ different $AR(1)$ models are statistically significant at standard thresholds. In addition to these rolling $AR(1)$ models, we have also experimented with the forecasts based on a set of $ARMA(1, 1)$ models estimated over the full in-sample period¹³.

¹³The $ARMA(1, 1)$ formulation is consistent with the assumption that the true latent loadings follow an $AR(1)$ model, but that the (observed) realized loadings are equal to the true loadings plus an independent measurement error.

The basic findings, available upon request, were very similar to the ones reported below for the $AR(1)$ models. Meanwhile, more complicated dynamic time series models, possibly allowing for asymmetries in the relationship between the past returns and the future factor loadings, may give rise to even better predictions¹⁴. To keep the comparisons simple, we do not purpose any of these more complicated model specifications here. Of course, the relatively short 5-year time span of high-frequency data also limits the scope of such investigations.

Table 3.5. Factor Loading Forecasts

	Constant	Rolling Regression	High Frequency Random Walk	High Frequency AutoRegressive
Market				
Bias	0.03	-0.68	0.00	0.01
MSE	0.01	0.49	0.01	0.05
MAE	0.07	0.68	0.09	0.04
Small Minus Big				
Bias	-0.52	-0.08	0.00	0.00
MSE	0.28	0.12	0.02	0.04
MAE	0.52	0.26	0.11	0.10
High Minus Low				
Bias	-0.03	0.94	0.01	0.00
MSE	0.03	1.02	0.06	0.03
MAE	0.13	0.94	0.18	0.03

Source: Euronext. The table reports the average forecast errors for forecasting the monthly realized factor loadings across the 9 portfolios and the 5-year out-of-sample period from January 2002 through December 2006. The constant predictions takes the factor loadings to be constant and equal to the conventional 5-year monthly regression based estimates at the beginning of the sample. Rolling Regression refers to the 5-year rolling regression estimates. High-frequency Random Walk assumes that the high-frequency-based realized factor loadings follow a random walk, or martingale, model. The High-frequency AutoRegressive predictions are based on a set of rolling $AR(1)$ models estimated over the previous 24 months. The rows labeled Bias, MSE, and MAE give the average errors, the average squared errors, and the average absolute errors, respectively.

Table 3.5 reports the average bias, Mean Squared Error (MSE), and Mean Absolute

¹⁴Braun *et al.* (1995) and Cho and Engle (1999) have documented statistically significant temporal variation and important asymmetries in CAPM betas, while Bollerslev and Zhang (2003), provide complementary evidence for asymmetric market risk exposure based on the same type of high-frequency measures utilized here.

Error (MAE), using each of the four procedures for forecasting the monthly realized loadings. The errors are averaged across the 9 size and book-to-market sorted portfolios and the 60-months out-of-sample period. The loadings associated with the market return, are generally the most accurate. Not surprisingly, the constant factor loadings (CON) result in the largest average Biases, MSEs and MAEs of all the procedures. The MSEs and MAEs for the RR and HF-RW procedures are fairly close. The one-step-ahead $AR(1)$ forecasts systematically result in the lowest MSEs and MAEs for all of the factor loadings. Intuitively, the HF-AR-based procedure more swiftly adjusts to changes in the underlying market conditions.

Tests of Pricing Performance

In order to focus directly on the potential benefits of using the high-frequency data in measuring and modeling the realized factor loadings, we rely on the actual realized excess factor returns when calculating the asset prices¹⁵. The three-factor Fama-French model has been the subject of numerous empirical studies and, as discussed above, has been found to perform very well in pricing the 9 size and book-to-market portfolios analyzed here. Indeed, for none of the different factor loading forecast procedures is the hypothesis of a zero “alpha” in the three-factor model is rejected by a simple t -test. Specifically, let the forecast error for the returns on portfolio i for month t based on the loadings from procedure m , be denoted by $e_{i,t}(m)$. The four t -statistics¹⁶ for testing the hypotheses that the average pricing errors across the 9 portfolios and the 60-month out-of-sample period are equal to zero *i.e.*, $\sum_i \sum_t e_{i,t}(m) = 0$.

Meanwhile, the superiority of the high frequency-based loadings is underscored by

¹⁵Since the errors associated with the forecasts of the factor loadings and the risk premiums are likely to be correlated, a rigorous true out-of-sample forecast comparison would also necessitate the joint modeling of the loadings and the factor risk premiums.

¹⁶The four statistics correspond to the four alternative predictions models for the factors loadings (CON, RR, HF-RW and HF-AR).

the results in Table 3.6, and the average absolute pricing errors, $|e_{i,t}(m)|$, reported in the first panel. Of course, the most appropriate statistical loss function for assessing the superiority of the competing forecasts will ultimately depend upon the application (see e.g., Christoffersen and Diebold, 1997 for a discussion of different loss functions in the comparison of nonlinear forecasting models). Hence, following Andersen *et al.* (1999), Table 3.6 also reports the heteroskedasticity adjusted average absolute standardized errors, $|e_{i,t}(m)/\sigma_{i,t}|$, and Table 3.7 reports the *pseudo*- R^2 defined by $1 - [e_{i,t}(m)/\sigma_{i,t}]^2$, where $\sigma_{i,t}$ refers to the realized volatility for portfolio i for month t constructed by the summation of the 5-minutes squared returns within the month augmented by the 1-day lead-lag autocovariances to account for the nonsynchronous trading effects. For the sake of parsimony, we only report the average loss across the 9 portfolios. From the results, reported in the first rows in each of the three panels, it is obvious that the HF-AR-based forecasts systematically yield the lowest absolute errors and the highest *pseudo*- R^2 's amongst the four different procedures¹⁷.

To formally test for the statistical significance of these apparent differences, Table 3.6 complements the average loss with the values of the three relatively simple pair-wise forecast comparison tests discussed by Diebold and Mariano (1995). In particular, let $d_{i,t}(m, n)$ denote the loss differential between the forecasts based on method m and n for portfolio i and month t . The $Z1_{(m,n)}$ statistic then refers to the standard t -test for the hypothesis that the mean differential is equal to zero. Alternatively, the null hypothesis of equal forecast accuracy, may also be tested by the nonparametric sign-test ($Z2_{(m,n)}$),

$$Z2_{(m,n)} = \left[\sum_i \sum_t 1I_{d_{i,t}(m,n)} - 0.5T \right] [0.25T]^{1/2}, \quad (3.36)$$

where $1I$ denotes the indicator function and $T = 60 \times 9$ refers to the sample size and m

¹⁷Although the improvements may seem numerically small, the increase in the R^2 for the HF-AR procedure is actually somewhat higher than the improvements reported in the study by Ferson and Harvey (1999) and Bollerslev and Zhang (2003) based on explicitly modeling the temporal variation in the monthly factor loadings using exogenous explanatory variables.

and n denote the monthly predictions models (CON, RR, HF-RW or HF-AR). Lastly, the related studentized version of the Wilcoxon's signed-rank test is formally defined by,

$$Z3_{(m,n)} = \left[\sum_i \sum_t 1_{d_{i,t}(m,n)} \times \text{rank}(|d_{i,t}(m,n)|) - 0.5T(T+1) \right] \times [T(T+1)(2T+1)/24]^{-1/2} \quad (3.37)$$

Under the null hypothesis that models m and n have the same predictive ability, all three test statistics should be asymptotically standard normally distributed. The values reported in the three lower panels in the table soundly reject the hypothesis of no superior predictive ability in favor of the high-frequency-based HF-AR procedure¹⁸.

¹⁸The results reported in Table 3.6 are based on the 60-month out-of-sample forecast comparison period and the simple $AR(1)$ model.

Table 3.6. Average Pricing Errors and Significance Tests

		CON	RR	HF-RW	HF-AR
	Panel A: Absolute Errors	0.1	0.08	-0.41	-0.01
$Z1_{(.,.)}$	CON		-31.76*	23.64*	-9.83*
	RR			20.35*	1.80
	HF-RW				-18.64*
$Z2_{(.,.)}$	CON		-25.62*	14.85*	-18.30*
	RR			12.82*	20.08*
	HF-RW				-131.98*
$Z3_{(.,.)}$	CON		-14.63*	-12.92*	-15.56*
	RR			-13.12*	-14.28*
	HF-RW				-15.67*
	Panel B: Absolute Standartized Errors	0.51	0.19	0.41	0.16
$Z1_{(.,.)}$	CON		-10.73*	-20.65*	-85.49*
	RR			-11.04*	7.32*
	HF-RW				11.58*
$Z2_{(.,.)}$	CON		-114.54*	-146.62*	-113.64*
	RR			-112.71*	54.99*
	HF-RW				116.37*
$Z3_{(.,.)}$	CON		-15.50*	-15.81*	-16.49*
	RR			-15.48*	-14.32*
	HF-RW				-13.24*

Source: Euronext. Panel A reports the average absolute errors. The entries in the first row have been multiplied by 10^2 . Panel B gives the average absolute standardized errors obtained by standardizing each error by realized volatility for that particular portfolio and month. All of the averages are calculated over the 9 portfolios and the 5-year out-of-sample period from January 2002 through December 2006. The procedure labeled CON takes the factor loadings to be constant and equal to the conventional monthly regression estimates at the beginning of the sample. RR refers to the 5-year rolling regression estimates. HF-RW assumes that the high-frequency-based realized factor loadings follow a random walk, or martingale, model. The HF-AR predictions are based on a set of rolling $AR(1)$ models estimated over the previous 24 months. $Z1_{(m,n)}$ refers to the t -test for a difference in the means across the monthly predictions for models m and n . $Z2_{(m,n)}$ gives the sign test while $Z3_{(m,n)}$ refers to the Wilcoxon's signed rank test for the same hypothesis. Under the null hypothesis of no superior predictive ability by either of the two models, all of the test statistics should be standard normally distributed. The sign * Indicates the significance at 5% threshold.

Table 3.7. Average of *Pseudo-R*² and Significance Tests

	<i>Pseudo-R</i> ²	CON	RR	HF-RW	HF-AR
		0.39	0.40	0.82	0.97
$Z1_{(.,.)}$	CON		70.81*	103.17*	58.31*
	RR			70.19*	-2.80*
	HF-RW				-74.95*
$Z2_{(.,.)}$	CON		114.54*	146.62*	113.67*
	RR			112.71*	-54.99*
	HF-RW				-116.37*
$Z3_{(.,.)}$	CON		-13.25*	-12.94*	-13.26*
	RR			-13.27*	-14.43*
	HF-RW				-15.52*

Source: Euronext. Table reports the average *pseudo-R*² defined by one minus the squared standardized error. All of the averages are calculated over the 9 portfolios and the 5-year out-of-sample period from January 2002 through December 2006. The procedure labeled CON takes the factor loadings to be constant and equal to the conventional monthly regression estimates at the beginning of the sample. RR refers to the 5-year rolling regression estimates. HF-RW assumes that the high-frequency-based realized factor loadings follow a random walk, or martingale, model. The HF-AR predictions are based on a set of rolling *AR*(1) models estimated over the previous 24 months. $Z1_{(m,n)}$ refers to the *t*-test for a difference in the means across the 9×60 monthly predictions for models *m* and *n*. $Z2_{(m,n)}$ gives the sign test while $Z3_{(m,n)}$ refers to the Wilcoxon's signed rank test for the same hypothesis. Under the null hypothesis of no superior predictive ability by either of the two models, all of the test statistics should be standard normally distributed. The sign * Indicates the significance at 5% threshold.

In order to further explore the performance of the HF-AR procedure, Table 3.8 reports the average pseudo *R*²'s for each of the 9 portfolios. Comparing the results to the *R*²'s associated with the conventional monthly rolling regression-based estimates, it follows from the second panel in the table, that the HF-AR procedure does better in an absolute sense. Interestingly however, for none of the portfolios where the RR procedure does better, is the difference statistically significant at the usual 5% level, whereas the HF-AR-based loadings result in statistically significant improvements for seven of the portfolios. Thus, in spite of the relatively short 5-year out-of-sample forecast comparison period, our various statistically tests systematically favor the asset pricing predictions based on the high-frequency factor loadings.

Table 3.8. Pseudo R^2 and Significance Test of Portfolios

	Low	Middle	High
Sample Mean			
Small	0.70	0.48	0.80
Middle	0.90	0.30	0.81
Big	0.91	0.85	0.89
$Z1_{RR,HF-AR}$ Test			
Small	-5.30*	-1.42	-2.55*
Middle	-3.73*	-0.88	2.67*
Big	6.78*	4.35*	3.68*

Source: Euronext. The first panel reports the *pseudo- R^2* for each of the 9 portfolios defined by one minus the squared standardized error for the HF-AR predictions averaged across the 5-year out-of-sample period from January 2002 through December 2006. The HF-AR predictions are based on a set of rolling $AR(1)$ models estimated over the previous 24 months. The second panel reports the t -statistic for testing identical mean R^2 's between the 5-year rolling regression estimates, RR, and HF-AR based predictions. Under the null hypothesis of no superior predictive ability by either of the models, the test statistics should be standard normally distributed. * Indicates the significance at 5% level.

Economic Significance of High-frequency Loadings

To highlight the economic value of the high-frequency-based loadings and corresponding asset pricing predictions, this section compares the *ex post* performance of the mean-variance efficient portfolios constructed from the 9 size and book-to-market sorted portfolios and the same four forecasting procedures discussed above. We construct the minimum variance portfolio as well as the portfolio that minimizes variance given a set target return, which is denoted μ_p , allowing for monthly rebalancing. To be precise, we solve the following two optimization problems:

$$\min_{\omega_t} \omega_t' \Sigma_t \omega_t \quad (3.38)$$

$$\text{s.t. } \omega_t' \mu_t = \mu_p \text{ and } \omega_t' \iota = 1$$

where ω_t is the $(N \times 1)$ vector of portfolio weights, and ι denotes an $(N \times 1)$ vector of ones. In addition, μ_t is the $(N \times 1)$ vector with conditional expected returns for the individual stocks, that is $\mu_t = E[r_t | I_{t-1}]$, where I_{t-1} denotes the information set available

at the end of day $t - 1$. Similarly, Σ_t is the $(N \times N)$ conditional covariance matrix.

The solution is of the problem is:

$$\omega_t = \frac{\mu_p \Sigma_t^{-1} \mu_t}{\mu_t' \Sigma_t^{-1} \mu_t} \quad (3.39)$$

In order to focus the comparisons directly on the impact of the forecasts for the factor loadings, we assume that the covariance matrix for the 9 test portfolios underlying the mean-variance optimizations is the same across the different procedures, and equal to the realized covariance matrix for the 24-month pre-sample period from January 2002 through December 2003. Also, we explicitly exclude the risk-free asset from the portfolio optimizations.

The first panel in Table 3.9 reports the resulting average monthly realized returns over the *ex post* January 2004 through December 2006 sample period for the optimized portfolios with the monthly expected standard deviations indicated in the first column. For all of the monthly standard deviations ranging from 0.01 to 0.02, or 0.03 to 0.06 at an annual level, the *ex post* returns are the highest for the HF-AR-based factor loadings. With the exception of the low-risk portfolio with a monthly expected standard deviation of 0.01, all of the Sharpe ratios for the monthly *ex post* returns and standard deviations, reported in the second panel in Table 3.9, also favor the HF-AR procedure by a fairly large margin. The very high *ex post* Sharpe ratios reflect the unprecedented bull-run during the 5-year out-ofsample period. Although it would be unrealistic to expect these same high values to obtain over other time periods, the ordering among the different procedures are nonetheless suggestive.

To value the performance gains associated with a particular trading strategy, we use the utility-based measure developed in Flemin *et al.* (2001 and 2002). This measure is based on quadratic utility as an approximation of the investors true utility function and assumes that relative risk aversion (γ) is constant. Under these conditions, the average

Table 3.9. Mean-variance Efficient Portfolios

Standard Deviation	Constant	Rolling Regression	High-frequency Random Walk	High-frequency AutoRegressive
<i>Ex post</i> Returns				
0.01	0.01	0.04	0.02	0.09
0.02	0.01	0.04	0.03	0.05
0.03	0.04	0.06	0.04	0.07
0.04	0.05	0.07	0.04	0.09
0.05	0.07	0.08	0.05	0.10
0.06	0.08	0.12	0.06	0.13
Sharpe Ratios				
0.01	0.25	3.74	1.75	8.75
0.02	0.38	1.88	1.38	2.38
0.03	1.25	1.78	1.08	2.25
0.04	1.19	1.61	0.94	2.19
0.05	1.35	1.51	0.95	1.95
0.06	1.26	1.96	0.96	2.13

Source: Euronext. The table reports the average monthly *ex post* percentage returns and corresponding Sharpe ratios for the mean-variance efficient portfolios constructed from the 9 size and book-to-market sorted test portfolios over the 5-year out-of-sample period from January 2002 through December 2006. The expected monthly standard deviations reported in the first column are based on the pre-sample realized variance-covariance matrix constructed from 5-min returns from January 2002 through December 2003. The expected return calculations labeled Constant takes the factor loadings to be constant and equal to the conventional 5-year monthly regression-based estimates at the beginning of the sample. Rolling Regression refers to the 5-year rolling regression estimates. High-frequency Random Walk assumes that the high frequency-based realized factor loadings follow a random walk, or martingale, model. The High-frequency AutoRegressive predictions are based on a set of rolling $AR(1)$ models estimated over the previous 24 months.

realized utility for a given portfolio (\bar{U}_p) is:

$$\bar{U}_p(\cdot) = W_0 \left[\sum_{t=0}^{T-1} R_{p,t+1} - \frac{\gamma}{2(1+\gamma)} R_{p,t+1}^2 \right], \quad (3.40)$$

where W_0 is the initial wealth and $R_{p,t+1} = R_f + \omega'_t r_{t+1}$ are the returns on the portfolio (p). This provides a consistent estimator of the investors expected utility. In order to measure the value of switching from one trading strategy to another strategy, we equate

their average realized utilities,

$$\sum_{t=0}^{T-1} (R_{2,t+1} - \Delta) - \frac{\gamma}{2(1+\gamma)} (R_{2,t+1} - \Delta)^2 = \sum_{t=0}^{T-1} R_{1,t+1} - \frac{\gamma}{2(1+\gamma)} R_{1,t+1}^2, \quad (3.41)$$

where $R_{1,t+1}$ and $R_{2,t+1}$ are the returns for the two strategies. To equate the average utilities, we subtract a constant, Δ , from each of the returns for strategy 2. This represents the most that the investor could give up (*i.e.*, a performance fee) and still have the same expected utility as under strategy 1.

Table 3.10. Willingness-to-Pay for Different Strategies

Standard Deviation	$\gamma = 1$	$\gamma = 10$
0.01	0.004	0.006
0.02	0.006	0.009
0.03	0.008	0.011
0.04	0.009	0.014
0.05	0.010	0.018
0.06	0.015	0.023

Source: Euronext. The table reports the monthly fees (in percentage points) that an investor with a quadratic utility function and constant relative risk aversion (γ) would be willing to pay to switch from the 5-year rolling regression-based return forecasts (RR) to the high-frequency-based $AR(1)$ forecasts for the realized factor loadings (HF-AR) for the 9 size and book-to-market sorted test portfolios. The fees are based on the monthly returns for the 5-year out-of-sample period from January 2002 through December 2006, with the monthly expected standard deviations for both methods determined by the pre-sample realized covariance matrix from January 2002 through December 2003.

The Table 3.10 reports the monthly fees that an investor with a quadratic utility function and constant relative risk aversion equal to γ would be willing to pay to switch from the traditional RR 5-year monthly rolling regression-based forecasts to the high-frequency-based HF-AR forecasts for the 9 portfolios returns. The fees in the table are expressed in monthly percentage points and determined empirically as the value which equalizes the *ex post* utility for the two different procedures when evaluated on the basis of the monthly returns¹⁹. To facilitate comparisons, we again fix the expected covariance

¹⁹The fee Δ refer, in the table, to the *ex post* monthly returns for the mean-variance efficient portfolio with the same benchmark expected standard deviations and expected returns determined by the HF-AR and RR loading forecasts, respectively.

matrix for both procedures at the realized covariance matrix for the January 2002 through December 2003 pre-sample period. Fixing the expected standard deviation at 6%, an investor with constant aversion risk $\gamma = 10$ would be willing to pay 2.3% per month, to switch from the Rolling Regression strategy to the high-frequency autoregressive strategy monthly return-based forecasts.

3.5 Conclusion

Following Bollerslev and Zhang (2003) using “Realized” measures and co-measures of risk in some classical asset pricing models, this chapter extends their approach in an extended Realized Conditional Asset Pricing Model framework. These extensions have included third-moment asset pricing models (see Jurczenko and Maillet, 2006) and conditional asset pricing models. We also make use of several methodologies aiming to neutralize data measurement and model misspecification errors (see Ledoit and Wolf, 2003 and 2004), trying to properly deal with inter-relations between financial assets in term of returns (see Zellner, 1962).

For the empirical investigation, high frequency data for some stocks on the french stock market have been used. Firstly, this chapter shows how high-frequency intra-day transaction prices for the different portfolios and the underlying factor return lead to more effective measures and models of the realized co-variations. As in Bollerslev and Zhang (2003), our empirical analysis are focused on a 1-month return horizon and the Fama-French test portfolios. In the French stock market on a limited sample, cross-sectional tests of the single-factor and three-factor model of Fama and French (1993) have shown that systematic risk measured by the covariance (or beta) with the market and other classical factors does not satisfactorily explain the cross-sectional variation in expected excess returns. We show also that the unexplained component in expected excess returns does not vary across portfolios and it is modest in magnitude.

This finding implies that additional variables representing portfolios characteristics we consider have no explanatory power for expected excess returns when coskewness is

taken in account. This result cannot be obtained if coskewness is neglected. In addition to that, our results have implications for testing methodologies, since they show that neglecting coskewness risk can cause misleading inference. Indeed, we find that coskewness is positively correlated with size. This suggests that a possible justification for the anomalous explanatory power of size in the cross-section of expected returns, is that it proxies for omitted coskewness risk.

Some tracks of improvement are however possible. It would be interesting to formally evaluate the importance of the errors of measurement of the sensitivities to the factors in the context of realized co-variations for the CAPM. Bai *et al.* (2000), Barndorff-Nielsen and Shephard (2002a, b), Meddahi (2002) or recently Griffin and Oomen (2008), provide theoretical results for the case of realized volatility. Instantaneous volatility is then broken up into volatility and autocorrelation term. Griffin and Oomen (2008) show theoretically and empirically that the bias of autocorrelation decreases with the sampling rate of the prices to be close to zeros after 25 minutes. A similar work could thus be made on realized beta and realized skewness. In the same way, by using a model with independent factors, the various relations between risks factors could be simplified. Indeed, by extraction of the common factors by an Independent Component Analysis with noises filtering (Kopriva and Seršić, 2008), all the cumulants matrixes will be diagonal (because of property of independence of the factors). So the estimate of the parameters with higher moments will be able.

Lastly, inspired by work of Brock and Hommes (1998) or Malevergne and Sornette (2006), the generalization of the previous approaches within a heterogeneous framework - where market participants confront their price- could constitute a natural extension of this study. For that, we can consider several sources of heterogeneity. Indeed, we can suppose that the market participants have different points of view about anticipations of the prices (Brock and Hommes, 1998).

Appendix

Appendix 1: Assumptions and Properties of Realized Moments

Following Beine *et al.* (2004), we present here assumptions and properties of Realized Moments. When computing the realized moments for closing prices (24h GMT), daily returns are defined as: $r_t = \sum_{i=1}^K r_{i,t}$ where K is the number of intervals per day (e.g. 24) and $r_{i,t}$ is the i -th intra-day return of day t .

A1.1: Realized Volatility

Consider two measures of the daily volatility $V(r_t)$: $V_{1,t} \equiv r_t^2$; *i.e.* daily squared returns, and $V_{2,t} \equiv \sum_{i=1}^K r_{i,t}^2$; *i.e.* the so-called realized volatility.

Assumption A_1 : $r_{i,t}$ is a martingale difference sequence, *i.e.* $E(r_{i,t} | \Omega_{i-1,t}) = 0$, where $\Omega_{i-1,t}$ denotes a filtration (information set) including past information on r up to (and including) the point in time $i - 1$ on day t .

Implications of Assumption A_1 :

1. $E(r_{i,t}) = 0$;
2. $Cov(r_{i,t}, r_{j,t'} | \Omega_{\max(i-1, j-1), \max(t, t')}) = 0$, which implies $Cov(r_{i,t}, r_{j,t'}) = 0$;
3. $E(r_{i,t} r_{j,t'}^k | \Omega_{i-1,t}) = 0$ if $i > j$ and $t'_t, k \in 1, 2, \dots$.

The implications follow immediately when we use the law of iterated expectations. Note that Assumption A_1 does not imply that $E(r_{i,t} r_{j,t'}^k | \Omega_{j-1,t'}) = 0$ if $i \leq j$ and $t < t'$.

Proposition 1: Under A_1 , $V_{1,t}$ and $V_{2,t}$ are unbiased estimators of $V(r_t)$.

Proof.

$$E(V_{1,t}) = E(r_t^2) \underbrace{=}_{A_1} V(r_t),$$

where $\underbrace{=}_{A_1}$ indicates the implication of the assumption A_1 .

$$\begin{aligned}
V(r_t) &\stackrel{A_1}{=} E(r_t^2) \\
&= E \left[\left(\sum_{i=1}^K r_{i,t} \right)^2 \right] \\
&= E \left(\sum_{i=1}^K r_{i,t}^2 + 2 \sum_{i=1}^{K-1} \sum_{j=1}^K r_{i,t} r_{j,t} \right) \\
&\stackrel{A_1}{=} E \left(\sum_{i=1}^K r_{i,t}^2 \right) \\
&= E(V_{2,t}).
\end{aligned}$$

Assumption A_2 : $E(r_{i,t} r_{j,t}^3 | \Omega_{j-1,t}) = 0, \forall i < j$.

This assumption imply the zero skewness and symmetry of the same conditional distributions.

Proposition 2: Under $A_1 - A_2$, $V_{2,t}$ is a more efficient estimate of $V(r_t)$ than $V_{1,t}$.

Proof. For the sake of simplicity, let us consider the case where $K = 2$.

$$\begin{aligned}
E(V_{1,t}^2) &= E[(r_{1,t} + r_{2,t})^4] \\
&= E(r_{1,t}^4 + r_{2,t}^4 + 4r_{1,t}^3 r_{2,t} + 4r_{1,t} r_{2,t}^3 + 6r_{1,t}^2 r_{2,t}^2) \\
&\stackrel{A_1-A_2}{=} E(r_{1,t}^4 + r_{2,t}^4 + 6r_{1,t}^2 r_{2,t}^2)
\end{aligned}$$

More generally, it follows that:

$$E(V_{1,t}^2) = E \left(\sum_{i=1}^K r_{i,t}^4 + 3 \sum_{i=1}^K \sum_{j \neq i}^K r_{i,t}^2 r_{j,t}^2 \right)$$

A similarly, for $V_{2,t}$,

$$\begin{aligned}
E(V_{2,t}^2) &= E[(r_{1,t}^2 + r_{2,t}^2)^2] \\
&= E(r_{1,t}^4 + r_{2,t}^4 + 2r_{1,t}^2 r_{2,t}^2),
\end{aligned}$$

or more generally:

$$E(V_{2,t}^2) = E \left(\sum_{i=1}^K r_{i,t}^4 + \sum_{i=1}^K \sum_{j \neq i}^K r_{i,t}^2 r_{j,t}^2 \right)$$

Since $(r_{i,t}^2 r_{j,t}^2) \geq 0$, it follows directly that $E(V_{1,t}^2) \geq E(V_{2,t}^2)$ and since $E(V_{1,t}) = E(V_{2,t}) = V(r_t)$, $V(V_{1,t}) > V(V_{2,t})$.

A1.2: Realized Skewness

Recall that the skewness is defined as: $E\{[r_t - E(r_t)]^3\}E(r_t^2)^{-3/2} \underbrace{=}_{A_1} E(r_t^2)E(r_t^3)^{-3/2}$. For this reason, a first measure of the daily cube returns are naturally $S_{1,t} \equiv r_t^3$. Extending the idea of realized volatility to the third moment would suggest the estimator $S_{2,t} \equiv \sum_{i=1}^K r_{i,t}^3$ is an unbiased estimator of $E(r_t^3)$.

Assumption A_3 : $E(r_{i,t}^2 r_{j,t} | \Omega_{j-1,t}) = 0, \forall i > j$.

Proposition 3: Under A_1, A_3 , $S_{2,t}$ is an unbiased estimator of $E(r_t^3)$.

Proof. For the sake of simplicity, let us consider the case where $K = 2$.

$$\begin{aligned} E(r_t^3) &= E[(r_{1,t} + r_{2,t})^3] \\ &= E(r_{1,t}^3 + r_{2,t}^3 + 3r_{1,t}^2 r_{2,t} + 3r_{1,t} r_{2,t}^2) \\ &\underbrace{=}_{A_1, A_3} E(r_{1,t}^3 + r_{2,t}^3) \\ &= E(S_{2,t}). \end{aligned}$$

The better estimators dominates in terms of efficiency depends on sign and size of higher moments as the simple case when $K = 2$ shows.

Proof.

$$\begin{aligned} E(S_{1,t}^2) &\underbrace{=}_{A_1} E[(r_{1,t} + r_{2,t})^6] \\ &= E(r_{1,t}^6 + r_{2,t}^6 + 6r_{1,t}^5 r_{2,t} + 15r_{1,t}^4 r_{2,t}^2 + 20r_{1,t}^3 r_{2,t}^3 + 15r_{1,t}^2 r_{2,t}^4 + 6r_{1,t} r_{2,t}^5 + r_{2,t}^6) \\ E(S_{2,t}^2) &\underbrace{=}_{A_1, A_3} E[(r_{1,t}^3 + r_{2,t}^3)^2] \\ &= E(r_{1,t}^6 + r_{2,t}^6 + 2r_{1,t}^3 r_{2,t}^3) \\ E(S_{1,t}^2) - E(S_{2,t}^2) &\underbrace{=}_{A_1, A_3} E(18r_{1,t}^3 r_{2,t}^3) + E(6r_{1,t}^5 r_{2,t}) + E(15r_{1,t}^2 r_{2,t}^4) + E(15r_{1,t}^4 r_{2,t}^2) \\ &> 0 \end{aligned}$$

Assuming independence over time of the intradaily returns, which implies Assumption A_3 , and assuming a symmetric distribution around the mean of these returns, we get that $E(r_{1,t}^3 r_{2,t}^3) = 0$ and $E(r_{1,t}^5 r_{2,t}) = 0$. In that case, $E(S_{1,t}^2) - E(S_{2,t}^2) = E(15r_{1,t}^2 r_{2,t}^4) + E(15r_{1,t}^4 r_{2,t}^2) > 0$, which implies that $S_{2,t}$ is more efficient than $S_{1,t}$.

Appendix 2: List of Firms

This table gives list of the 43 firms included in the analysis. For these stocks, the database includes a complete set of five-minutes transactions prices from January 02, 2002, through December 29, 2006 (1,281 trading days).

Code	Names	Code	Names
3232	TF1	8058	SCHNEIDER ELECTRIC
3340	EADS	8088	VIVENDI ENVIRON.
7886	AIR LIQUIDE	8131	SAINT-GOBAIN
7894	CARREFOUR	8146	CAP GEMINI
7903	TOTAL FINA ELF	8155	CANAL +
7908	OREAL	8157	VINCI
7910	VALLOUREC	8159	CASINO GUICHARD
7914	ACCOR	8181	AGF
7924	BOUYGUES	8194	VIVENDI UNIVERSAL
7926	SUEZ	8217	DEXIA SICO.
7927	LAFARGE	8279	STMICROELEC.SICO.
7931	SANOFI SYNTHELABO	8287	ALCATEL A
7935	AXA	8301	LAGARDERE
7937	GROUPE DANONE	8312	VALEO
7942	PERNOD-RICARD	8335	PUBLICIS GROUPE
7971	LVMH MOET VUITTON	8350	SOCIETE GENERALE A
7990	SODEXHO ALLIANCE	8356	BNP PARIBAS
7994	MICHELIN	8421	RENAULT
8000	THALES	8546	FRANCE TELECOM
8013	PINAULT PRINTEMPS	11149	THOMSON
8015	PEUGEOT	84569	CREDIT AGRICOLE
8029	ESSILOR INTL		

Source: Euronext. This table reports a list of the 43 firms included in the analysis. For these stocks, the database includes a complete set of five-minutes transactions prices from January 02, 2002, through December 29, 2006 (1,281 trading days).

Chapter 4

Recovering an Intrinsic Well-behaved Business Time from Market Prices

4.1 Introduction

Asset prices move in response to the arrival of new informations. Generally speaking, some days are quiet - with no specific news impacting the price - whilst others are turbulent and full of informations. Ideally, we should then pay more attention when trading is intense and prices move much faster, than when it is slow and prices only fluctuate smoothly. Formally, Clark (1973) first introduced the idea that the shape of the distribution of financial asset returns might be due to a superposition of stochastic processes, the price process being subordinated to a general information process. In this general Mixture of Distributions Hypothesis (MDH) framework, the asset returns are governed by a first stochastic process - a classical diffusion model, itself driven by a variable representing the flow of informations on the market, and modeled as a second stochastic process. Volatility of returns is then random and classified as stochastic process due to the randomness of news. The new information defines indeed an operational time, or

business time, contrasting with the usual calendar-time measure, where observations are regularly spaced by the clock. In the business time, recorded prices are thus not anymore equally spaced in real time as daily, weekly or monthly series, but equally spaced in terms of information. The variable governing the second process itself shall then be a *proxy* for the flow of informations. In practice, volume has been widely used (see for instance Epps and Epps, 1976 and Harris, 1986), as well as the number of transactions or the volatility of the series itself when those are not available (*i.e.* the theta-time by Dacorogna *et alii*, 2001). Ané and Geman (2000) and Geman *et al.* (2002) show, both from a theoretical and practical point of views, that there is always some time deformation that will transform the price process in the classical Brownian motion with the right conditioning. These authors argue that the intensity (number of transactions) can play that role. These results, supported by Gabaix *et al.* (2003), have recently been intensively questioned and criticized by Guillemot *et al.* (2005), indicating as a result that the choice of the clock is not so evident. In other words, any business time is far from leading to homoskedastic Gaussian return series with no volatility clustering phenomenons.

The aim of this article is to recover some implicit time, linked to the trading activity and the information arrival, for obtaining rescaled financial series that are well-behaved - *i.e.* in conformity to traditional hypotheses in the mean-variance paradigm - for practical uses, such as asset pricing and allocation or risk management. Like when performing a Box-Cox transformation from non-normal to Gaussian series, we simply would like to transform the initial series into normal ones to fully exploit the classical tools of analysis. Put unlike a Box-Cox type of transformation however, all returns in the new rescaled series will correspond to true observed prices, measured by slowing or accelerating the clock according to the information flow. Our simple intuition is that it is possible to get improvements in financial applications by using true price series with good properties, simply by changing the time stamp of the quotes and avoiding the arbitrariness of relying on closing prices.

In the current article, we first begin to set up a simple procedure for deriving a special implicit business-time, starting from a time series of asset prices, and finally conclude by empirical applications. More precisely, the chapter is organized as follows. The first section describes the procedure, that yields a sampling scheme that produces returns as close as possible to the Gaussian distribution. The second section is an univariate application on the stocks composing the CAC40 index, taken at high frequency, stock by stock. The third section treats simultaneously all stocks, yielding one business scale for the whole sample such as the combined returns of all stocks are Gaussian. We use this last scale to derive practical applications, checking in particular the stability and values of correlations. Final section concludes.

4.2 Recovering an Intrinsic Time

Our goal is to build a simple procedure that will allow us to get business time prices from calendar time prices alone. This procedure should ideally take into input a time series of prices, and deliver a series of sampling times, necessarily at a lower frequency. The returns computed at those points will be as close as possible to the Gaussian distribution, as defined by an objective function. The optimization involved is thus pretty unusual, since the parameters to adjust are the timing of each sample, and not functional parameters. For this task, genetic algorithms (Holland, 1998) are well suited, since they can actually deal with fairly large problems, sometimes at the cost of lengthy converging times, and are designed to work with discrete parameters. The algorithm is stochastic, meaning that if several solutions exist that satisfy the termination *criteria*, any of them can be picked from the initial random pool - that is adequate for our purpose, but could prove problematic in practice, if unique results were needed.

The algorithm uses a population of possible solutions, encoded in vectors. Following the genetic analogy, each solution is called a chromosome and the current set of individuals is called the gene pool. Here, each chromosome is a set of increasing numbers representing

the number of observations in the original series, *i.e.* a point in time (day, hour and minutes). The length of each genome is the number of observations desired in the new business-time series. The size of the population is a compromise between the speed of execution and the minimal breadth of solutions needed to maintain a pool of competing solutions. In this application, population is set at 300.

In the current application, the genomes' lengths were fixed so that the number of observations is equal to the number of business days on the Euronext market on the considered period. From the 2nd of January 2002 to the 29th of December 2006, the sample consists of 1,281 days, *i.e.* on average 1.4 calendar day between each closing price (when one includes week-ends, holidays and suspensions). Each gene actually encodes an integer, the position of the observation in the original 5 minutes sampled series, which translates as one of the dates for which a price is available (see for instance Table 4.1). The genome thus ultimately yields a series of prices, whose returns will be used as input of the fitness function.

Table 4.1. A Sample Genome (for the stock *TF1*)

Gene	Value	Price
1	01/02/2002 09:40	28.68
2	01/03/2002 12:30	29.68
3	01/08/2002 14:20	30.10
4	01/09/2002 14:05	30.70
5	01/09/2002 17:35	31.00

Source: Euronext, 5' intraday CAC40 quotes on the period 01/02/2002 to 12/29/2006. This example illustrates a one-point crossover operation

The objective function to be maximized is indeed called in our context the fitness function, and it will be used to rank the individuals in the population, from best fit to least fit. As there are various ways to test the normality of a set of observations, the choice of a specific objective function depends on the aim of the optimization. We use here two different *criteria*, either alone or in combination: the Jarque-Bera statistic and the Anderson-Darling statistic. The Jarque-Bera statistic is a combination of the squared

skewness and the squared excess kurtosis of the distribution, and follows a Chi-Square distribution under the null hypothesis of normality. The Anderson-Darling statistic is a goodness-of-fit measure that uses log-distribution functions, thus putting more emphasis on the tails; its distribution for the Gaussian case has been tabulated.

Like all numerical optimization algorithms, genetic algorithms work iteratively, by generating increasingly better solutions until the termination conditions are met. Starting from a given population, an iteration creates a new population of the same size, called generation. Genetic operators are applied to the parent population to create the descendant. Accordingly, the first step is then parents' selection, then choice of the operators to apply to it.

Five main procedures exist to select parents: the Best method, the Tournament method, the Fitness method, the Fitness-overselection method and the Random method. In each case, the method has to balance the twin aims of picking the best individuals (to improve the objective function) and maintaining diversity in the population (to avoid getting stuck in a local *extremum*).

In the Random method, parent individuals are selected independently of their fitness score, thus yielding a pure random drift across the space of possible solutions. In the Fitness method the individuals which will breed the next generation are still selected randomly, but weighted by their fitness, thus yielding a directed random walk. In the Tournament method, the best individuals in a random sample are kept, whereas in the Fitness-over selection method, individuals are thus grouped according to their fitness score, before being selected with a specific probability for each group. We use here the Best method, in which the best individuals, ranked by fitness, are used to generate the next generation.

Though the results should be robust to the choice of the algorithm, Chidambaran

et al. (1999) report significant differences in convergence speed across the methods, to be balanced with the computational times required for the processing of each generation. The space of solutions here is both of very high dimension (1,281 different genes, or sampling points) and heavily constrained (the genes are arranged in monotonously increasing order). Since, the efficiency of the algorithm will not only depend upon the choice of the parent population, but of the method used to produce the next generation; we will detail these methods below, starting with operation of mutation.

The mutation operator is, strictly speaking, the only one needed for convergence. It changes randomly the genome of part of the population according to pre-defined rules. The mutation operator should not change too much the best individuals (in order not to reverse the improvements of the algorithm), make slight changes (in order to get marginally better solutions) and bring new solutions in the population (in order to expand the solution space beyond the combinations of the initial population). Mutation operator can be purely deterministic or stochastic in the sense that changes arise with respect to a density law. In our case, the mutation strikes randomly in the population, sparing the best individuals (see Table 4.2). The mutation operator will simply change one bit in the genome, replacing one of the dates by any other between the previous and next genes.

Table 4.2. Mutation of a Genome

Gene	Parent (Before Mutation)	Descendant (After Mutation)
1	01/02/2002 09:40	01/02/2002 09:40
2	01/03/2002 12:30	01/03/2002 12:30
3	01/08/2002 14:20	01/03/2002 17:25
4	01/09/2002 14:05	01/09/2002 14:05
5	01/09/2002 17:35	01/09/2002 17:35
⋮	⋮	⋮

Source: Euronext, 5' intraday CAC40 quotes on the period 01/02/2002 to 12/29/2006. This example illustrates a mutation operation. Only the third gene changes here.

As for the selection of the parent populations, there are multiple procedures to choose from for the reproduction operator. The aim is to breed new individuals from the selected

population of best solutions. The simplest way to proceed would be to simply duplicate in the new population the parents' genomes as they are. This solution, however, relies only on the mutation to bring new solutions in the gene pool, and thus relatively inefficient. A better way is to combine characteristics of two parents, in a process analogous to sexual reproduction. This technique, known as crossover, consists in taking the genome of one of the parents, and swap part of it with the corresponding genes of the other parent. In the simple one-point crossover that we use here, a single pivot is selected in the genome, at a random point among the 1,281 genes. The two parents' genes are compared at this point, and the new individual's gene consists in the first parent's gene up to the pivot, and the second parent's gene from the pivot to the end. The roles of the first and second parents are swapped as needed to insure that the child is a viable individual, *i.e.* its genome consists in an increasing series of dates (Table 4.3).

Table 4.3. One Point Cross-over

First Parent Genome		Second Parent Genome		Descendant Genome
⋮		⋮		⋮
03/08/2002 13:10		03/11/2002 10:00		03/08/2002 13:10
03/11/2002 15:15		03/11/2002 16:15		03/11/2002 15:15
03/13/2002 14:25	+	03/13/2002 15:45	=	03/13/2002 15:45
03/14/2002 10:00		03/19/2002 14:15		03/19/2002 14:15
03/15/2002 12:20		03/22/2002 09:15		03/22/2002 09:15
⋮		⋮		⋮

Source: Euronext, 5' intraday CAC40 quotes on the period 01/02/2002 to 12/29/2006. This example illustrates a one-point crossover operation.

In parallel with the creation of new solutions, the algorithm should deal with the bad individuals in the population. Though they do not contribute to the immediate improvement of the objective function, it might be desirable to keep them alive in order to maintain diversity in the gene pool. We choose here to keep the population size constant, and thus, after determining the best individuals, we use their children to replace the genome of the worst performing individuals.

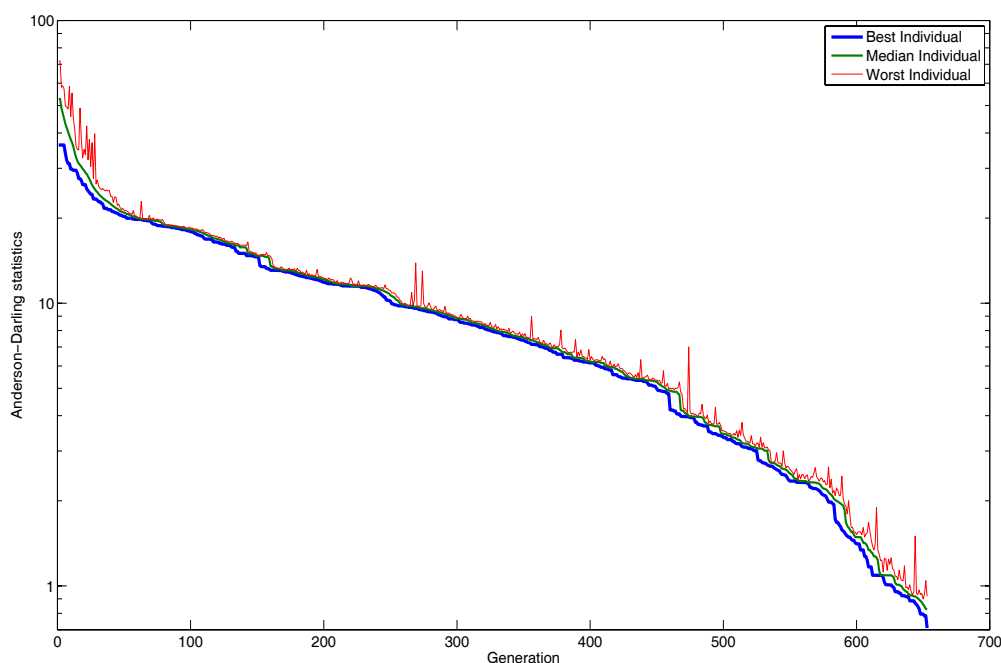
Starting from a parent generation, the new generation will consist of the best individuals of the parents population, unchanged, mid-performing individuals of the parent generation subjected to mutation and a cross-over of the best individuals to replace the worse performing ones.

In order to the convergence of the algorithm, there has to be one or several solutions satisfying the termination conditions, *i.e.* such as the Anderson-Darling normality test cannot reject the null hypothesis of a Gaussian distribution of the business-time returns. Ané and Geman (2000) argue that there is a business time in which the asset price process is a Brownian motion; it is less obvious in practice whether the discrete nature of the sampling, even in high frequency, insures that this is always the case. As the frequency increases, it is clear that it will become more and more difficult to find this business time, since the deviation from normality is higher to start, and the possible sampling schemes are less numerous. Though we present here the results for daily data, preliminary experiments suggest that finding a business time scale for higher frequencies is difficult, though still possible up to a frequency of two observations a day. The convergence can be speed up by several practical choices, notably an educated choice of starting pool. We have used here a totally random starting pool, but we are still exploring different initial conditions to improve the process.

In the case of daily data, there might be several solutions satisfying the constraints, and theory does not dictate the choice. Depending on the normality test used as fitness function, we will also find different solutions, since the exact conditions being fulfilled are not equivalent. We present in this chapter the results of convergence using the Anderson-Darling *criterion*. The following figure 4.1 shows the decrease of the Anderson-Darling statistics with each generation, starting from a random pool and ending with the business-time scale that we aimed for.

The convergence graph presents simultaneously the mean and best values of the statistics across the whole population. As the best solutions gradually diffuse to the rest of the population, the mean value gets closer to the *optimum*. The graphic is in log-scale, so

Figure 4.1. Convergence of the Anderson-Darling Objective Function



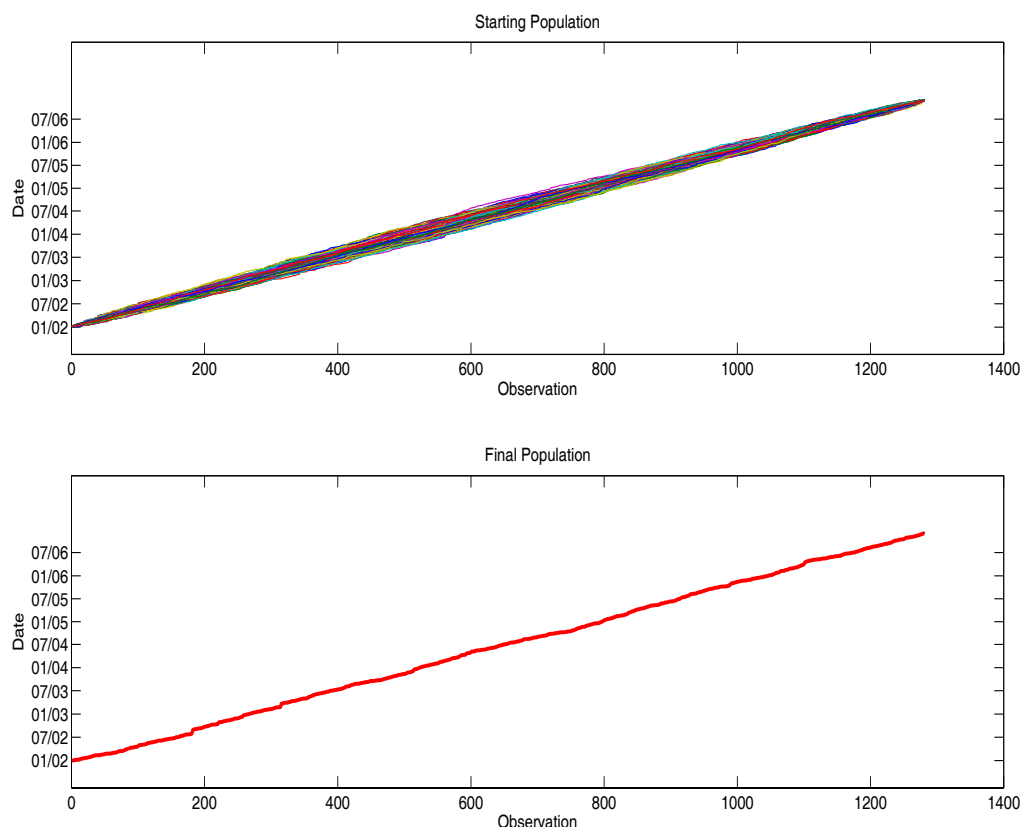
Source: Euronext, 5' intraday CAC40 quotes on the period 01/02/2002 to 12/29/2006. The Anderson-Darling statistics on business time returns are taken from 5 minutes prices. This figure illustrates the convergence of the Genetic Algorithm with the iterations of the procedure.

the slope indicates a regular rate of improvement of the statistics, though the initial rate is much higher, as it is in the final steps. Local convergences occur when a good solution diffuses across the gene pool, global convergence when it is replaced with a better one, though the process can take hundred generations.

The optimization process can be stopped according to several *criteria*, of increasing difficulty: when no further improvement is deemed possible (even if Gaussianity is not achieved), when one of the two statistics falls below the 5% rejection threshold, when both statistics fall below the rejection threshold. By its nature, the genetic algorithm relies on a diversified population to achieve optimization, thus a relatively large population is needed to avoid *de facto* termination when a single gene overrides all competitors. In practice, for all the 43 stocks, the termination *criterion* used was the Anderson-Darling

test. Trying to impose both normality tests together impedes convergence with the simple genetic algorithm we used here, but further extensions of these results will include tools tailored for multi-objective optimization in order to force convergence.

Figure 4.2. Convergence of the Solutions towards a Single Genome



Source: Euronext, 5' intraday CAC40 quotes on the period 01/02/2002 to 12/29/2006. The dispersion of the starting and final genomes, with a population of 300 individuals.

Figure 4.2 represents starting and ending points of the optimization. Each line is a single genome, *i.e.* a set of increasing numbers representing the position of each new point in the original series. The upper graph shows the range of the initial random population. The lower graph shows the convergence of the population to an optimum, all genomes nearly identical. As the best solution diffuses through the whole population, the diversity diminishes until all individuals map the same solution, or almost (the mutation operator insures that the population is never totally at rest).

4.3 Characteristics of the Intrinsic Business Times

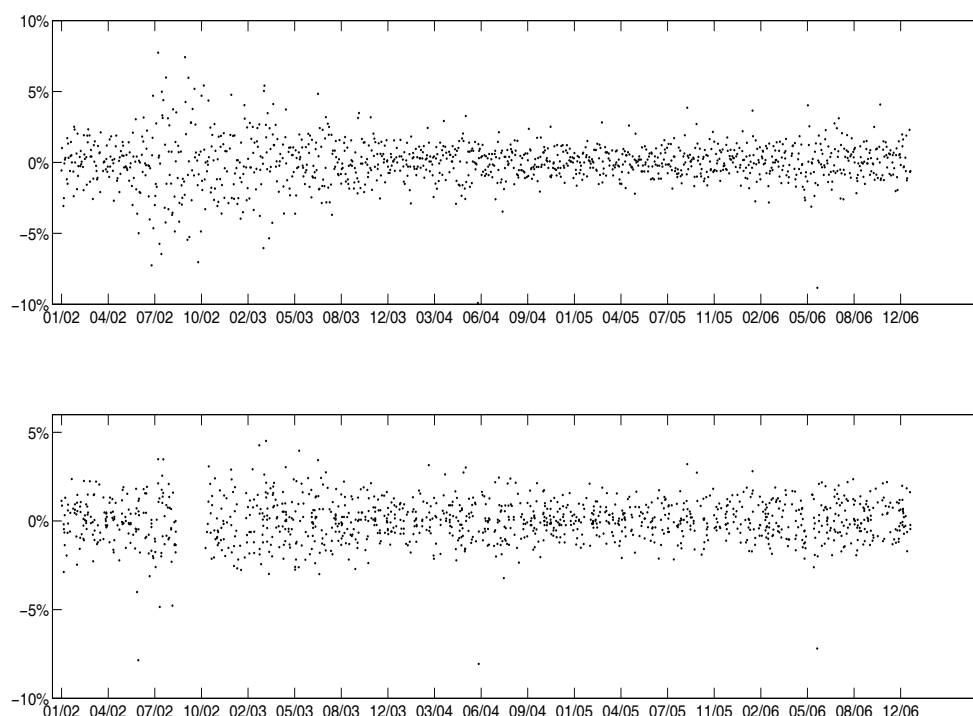
The data consists of 43 stocks, between the 2nd of January 2002 and the 29th of December 2006, sampled at a 5 minutes interval. Each five minutes, the last price for each stock was collected from *Euronext*. The number of observations *per* day is thus not perfectly regular, especially since *Euronext* opening hours have changed during the period (extended by half an hour).

Daily closing prices being the most common observations available, we decided here to use them as a benchmark, though we have experimented with higher frequencies as well. Thus, the business time will be defined as comprising the same number of observations as the number of trading days within the sample. All quantities will thus be on the same order of magnitude as daily data, and the average number of observations *per* trading day will be, by construction, equal to one.

The optimisation process was repeated on the 43 stocks, according to the Anderson-Darling termination *criterion* in the following illustrations. As explained above, the convergence is always achieved, though this is not the case when using both Anderson-Darling and Jarque-Bera *criteria*.

The example is typical of the effect of time deformation (see Figure 4.3. The tails are flattened, with both a decrease in the range of results and in the rate of occurrence of extreme results (Figure 4.4). The white spaces on the time axis for business returns indicate the periods where the sampling scheme skips prices which carry little information, such as for instance in May 2005. It should be noted that since, unlike an exchange rate, the time series record for a stock is not continuous, there are irreducible jumps in prices between closing time and the next opening time, which explain that some extreme returns remain, since they cannot be divided in shorter parts.

Even though the normality *criterion* in this case is Anderson-Darling's goodness-

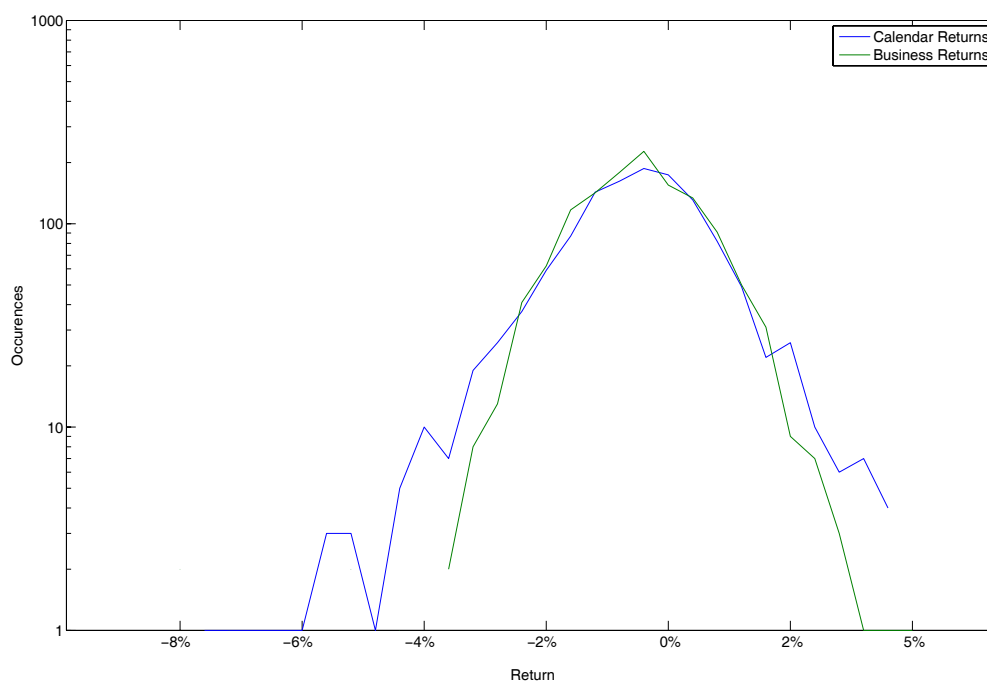
Figure 4.3. Calendar (top) and Business (bottom) Returns for *Air Liquide*

Source: Euronext, 5' intraday Air Liquide quotes on the period 01/02/2002 to 12/29/2006. Both calendar time and business time are represented on this figure.

of-fit tests, the time deformation generally reduces the amplitude of the skewness and decreases the *kurtosis* (Figure 4.5). The P-P plot (see Figure 4.6) shows, for Air Liquide, that the convergence under the Anderson Darling is obtained with the procedure. Since the business time returns distribution is here accepted to be normal, whereas the calendar returns exhibit the familiar deformation in the tails and wings of the distribution.

After verifying that the returns satisfy the normality condition, the properties of the time scale can then be studied for themselves. The normality of the returns has been achieved by compression and dilatation of the time, the patterns in the sampling shall then reflect the variations in the rate of informations imparted to the price.

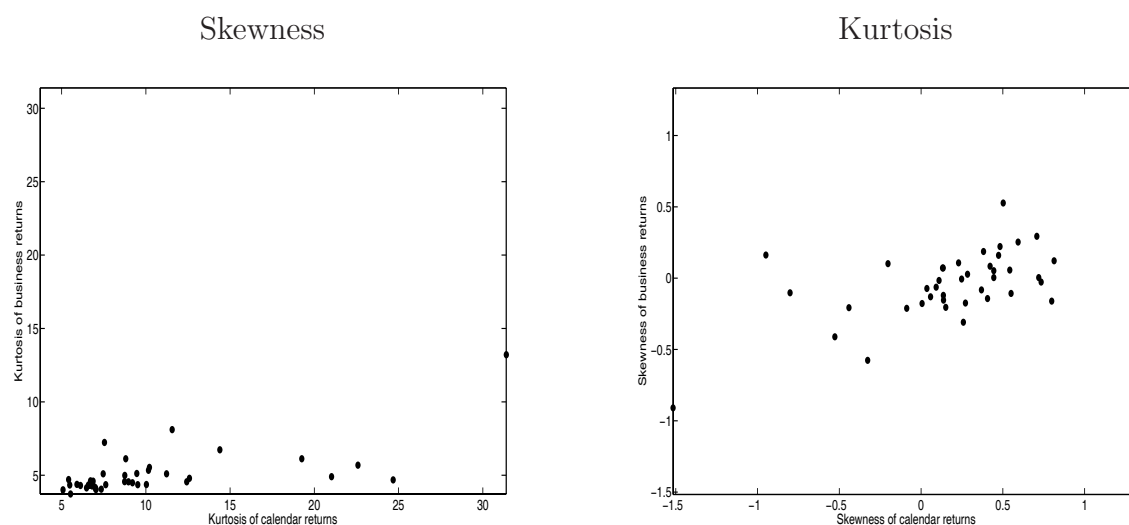
As could be expected, the business time picks up more observations when market moves are large, and less in quiet times. However, there is no clear indication of regime

Figure 4.4. Log-densities for *Air Liquide* calendar and business returns

Source: Euronext, 5' intraday Air Liquide quotes on the period 01/02/2002 to 12/29/2006. Both calendar time and business time are represented on this figure.

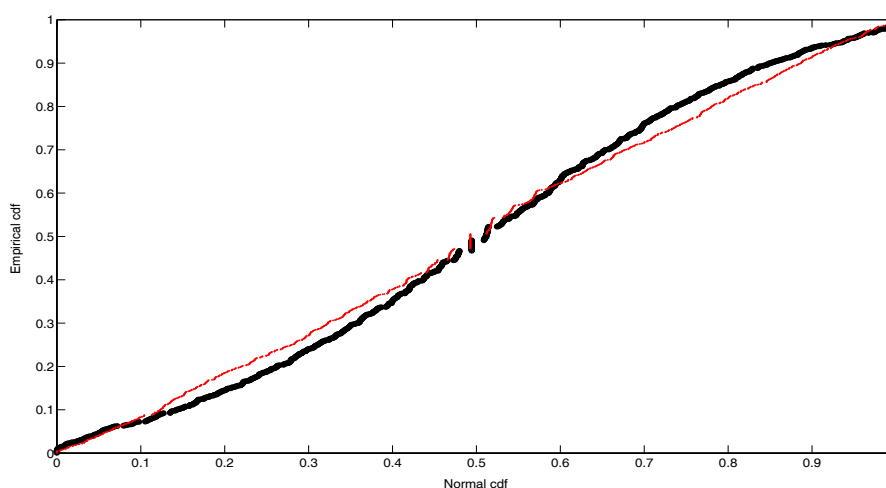
shifts in the price generating process. The days where more observations are taken tend to occur near each other (summer 2003 for instance), the persistence is not extremely marked, and there is not a division of the sample between a slow period and a fast one: sampling rate remains relatively uniform through the whole sample.

Figure 4.5. Moment-based Normality Diagnoses



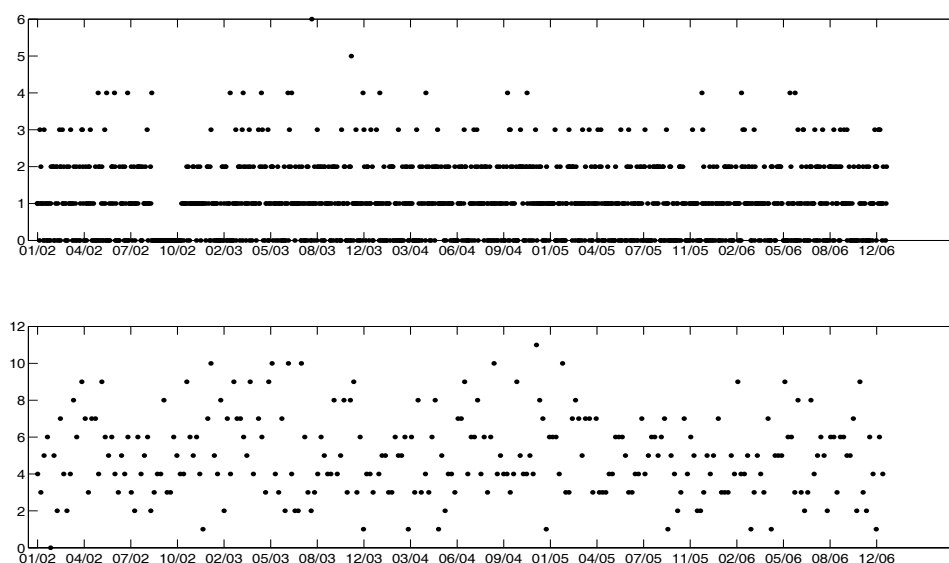
Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. Both calendar time and business time are represented on this figure. Each data point represents a couple of skewness and *kurtosis* of calendar and business returns in a specific time for each of the 43 sample stocks series.

Figure 4.6. P-P Plot Normality Diagnosis for Air Liquide Returns in Calendar Time (black) and Business Time (red)



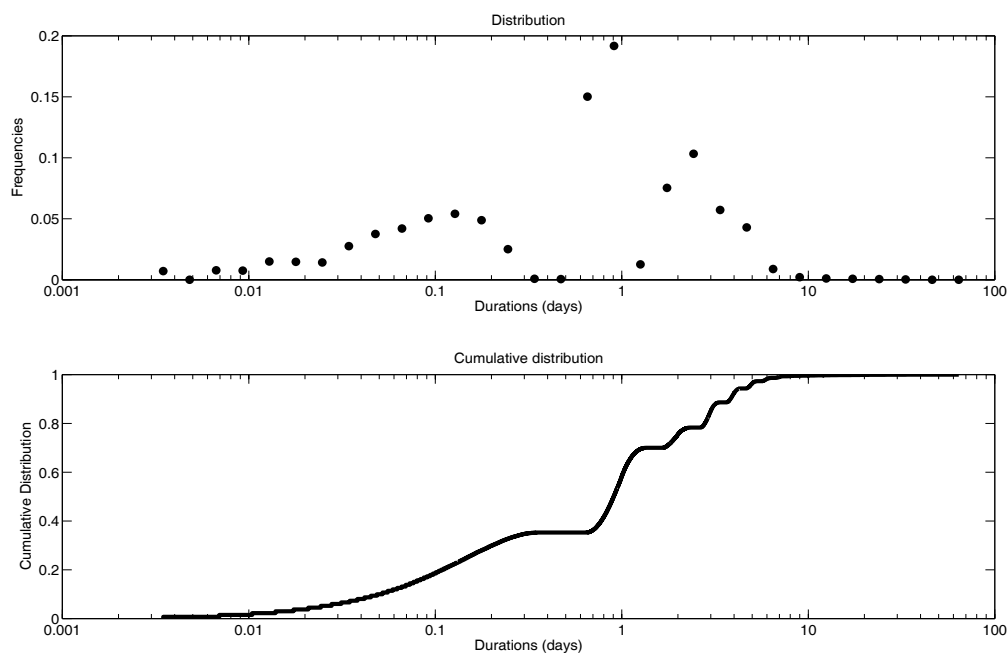
Source: Euronext, 5' intraday Air Liquide quotes on the period 01/02/2002 to 12/29/2006. Both calendar time and business time are represented on this figure.

Figure 4.7. Number of Daily (top) and Weekly (bottom) Sampled Observations for *Air Liquide*



Source: Euronext, 5' intraday Air Liquide quotes on the period 01/02/2002 to 12/29/2006. Both daily times and weekly times are represented on this figure.

Figure 4.8. Distribution of the Log-durations of the 43 Business-time Clocks



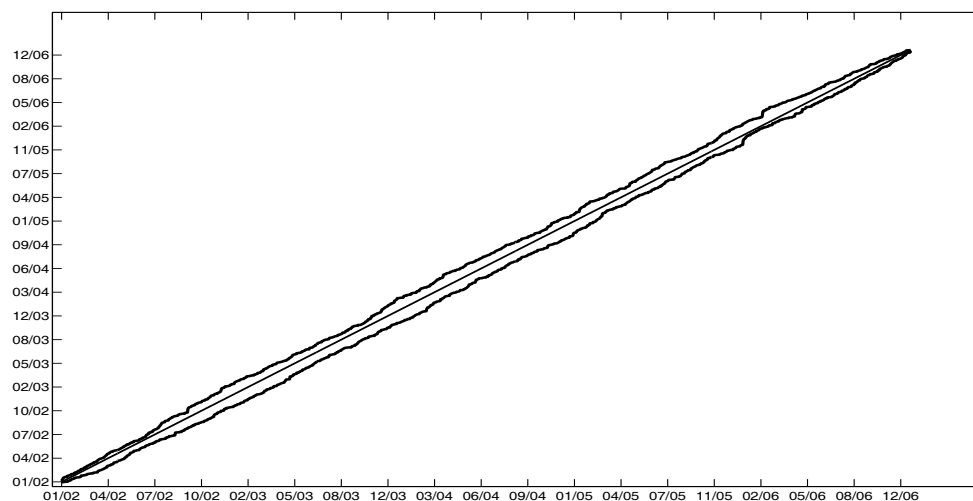
Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. The figure plots empirical frequencies and cumulative distribution of the Log-durations.

The reciprocal of the frequency is the duration, *i.e.* the time interval between two observations. The durations are distributed over a wide interval, the observations being sampled from 5 minutes apart (the minimum period possible) up to 78 days apart Figure 4.7. The distribution of business times is thus highly asymmetric, with a peak at one day. The discontinuities observed in the function correspond to the intervals for which no data exists, mainly the existence of an overnight period (Figure 4.8).

Each optimization yields a different business clock for each stock, with the range of possible values shown below (Figure 4.9). The x-axis represents the location of the observations in calendar time, the y-axis the location of the observations in each of the individual business time. We represent for each observation, the minimum and maximum of the values taken on stocks. The range of the bands is typically about three months between the slowest moving stock and the fastest moving one.

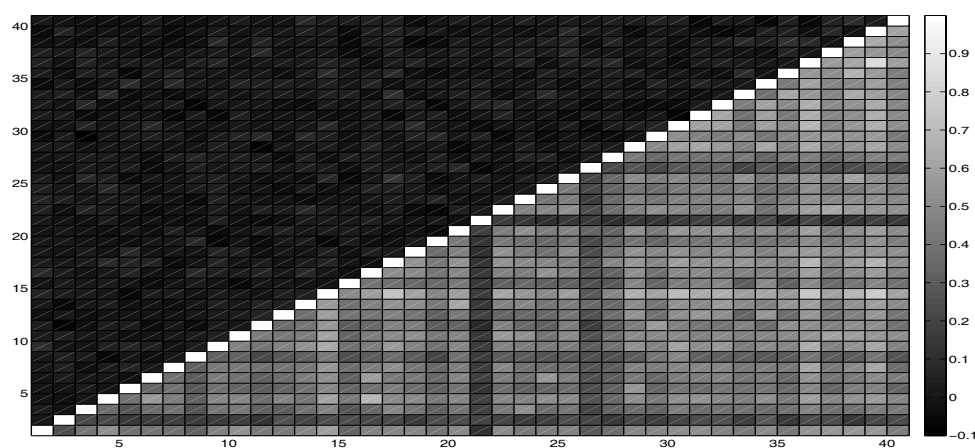
Likewise, the comparison between the correlations of the returns in calendar times and business time shows that the latter is tailored for each stock, and thus the resulting de-synchronisation implies that the correlations vanish (see Figure 4.10). None of the correlations in business time appears significant, since none of the stocks useS the same sampling scheme. These results show that for any multivariate application, a different sampling scheme is required, in order to select a business time common to all stocks under consideration.

Figure 4.9. Envelope of Business Clocks over the Sample of 43 Stocks



Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. The figure represents for each observation, the minimum and maximum of the values taken on stocks in the calendar time and in the business time.

Figure 4.10. Correlations in Calendar Time (bottom) and Business Time (top)

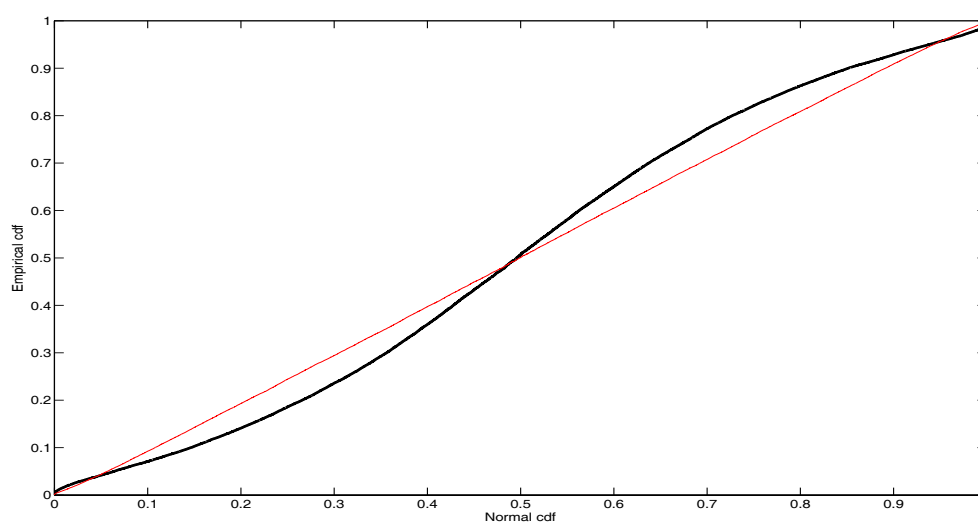


Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. Correlations between daily and business returns are represented here.

4.4 A Common Business Time for the Stock Market

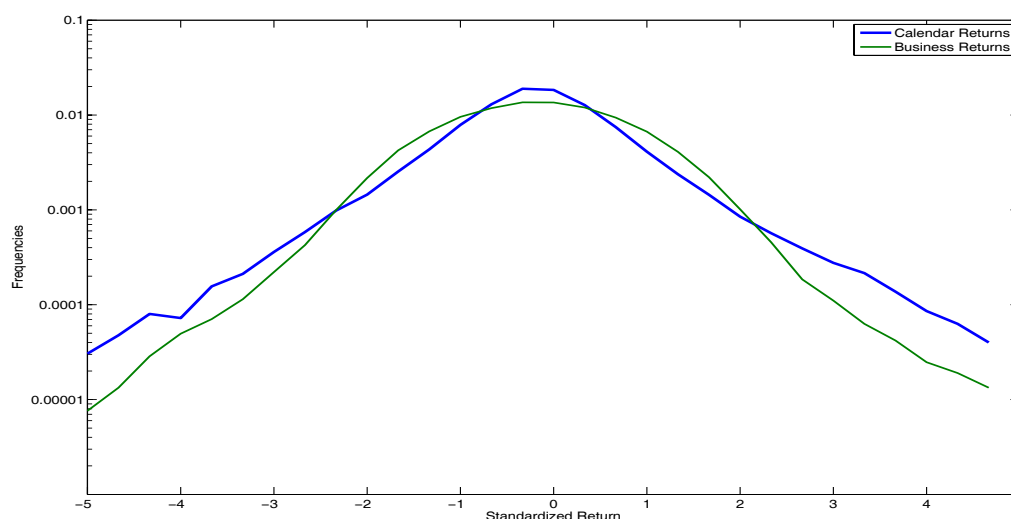
In order to exploit the regularity of business time returns in a multivariate framework, it proved necessary to have the same time-scale for all stocks, since no commonality emerges by itself from the individual optimizations. We will thus apply the same optimization technique as before, with the goal of getting a unique business time-scale for all stocks, thus putting all business time returns in the optimization function simultaneously. The process becomes much more constrained, since the normality has to be verified with the same sampling scheme for different stocks. We use hereafter the Anderson-Darling *criterion*, reasoning that if every one of the 43 stocks is a Gaussian variable, then the aggregation of standardized returns shall also be Gaussian, with zero mean and unitary variance. Though the Anderson-Darling *criterion*, on the data used, does not fall below the significance threshold, minimizing this statistic still deforms the shape of the returns towards the Gaussian distribution (see Figure 4.11).

Figure 4.11. P-P Plot of Standard Returns in Calendar (Black) and Business (Red) Times



Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. The figure plots the empirical cumulative distribution function against the normal cumulative distribution function for the standardized returns in calendar (black) and business (red) times.

Figure 4.12. Log-Densities of Standard Returns in Calendar and Business Times

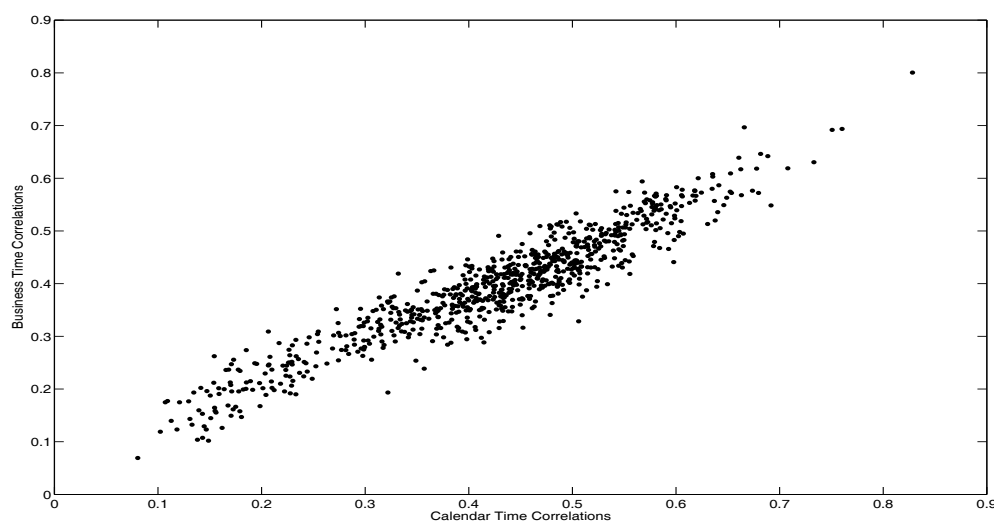


Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. The figure plots the Log-densities empirical distribution of the standardized returns in calendar and business times.

Likewise, the log-densities show the characteristic flattening of the tails and decrease of the kurtosis (Figure 4.12). We thus use the business time scale common to all stocks to compute the correlations and variances.

Figure 4.13 indicates that the imposition of a common time scales allows us to recover significant correlations. The pair-wise correlations in calendar and business time are strongly similar (the Spearman correlation coefficient is .9). Though the extreme correlations are identical, there are many individual differences in the middle range, meaning the actual weights of portfolios obtained from the matrix will be different.

A way to visualize the impact of the change in correlations is to use the correlation matrix as the basis of a distance matrix, and to then group stocks in clusters based on their distance. Figure 4.14 shows the two groupings, with calendar and business times' correlations. The algorithm used here puts groups (which can be on individual stock) together using the smallest distance in each iteration. The distance between two groups of stocks is defined as the lowest correlation between the members of the groups.

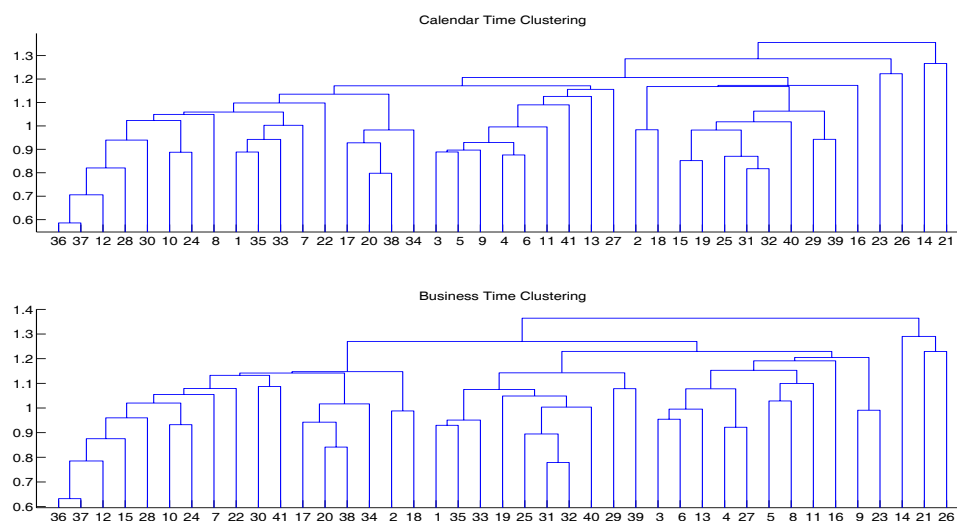
Figure 4.13. Pair-wise Correlations in Business Time *versus* Calendar Time

Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. The business-time correlations are plotted against their calendar-time values.

In both cases, the lowest level clustering broadly corresponds to sectors (financial, high tech, distribution), but the exact attachment of branches at higher levels differs across time scales. The usual test for variance reduction is to build the minimum variance portfolio and see how it fares out-of-sample. We conduct here this test by using 260 rolling observations (one year in calendar time) to build variance-covariance matrixes in business and calendar time, which yield for each period a new minimum variance portfolio. The return of this portfolio is then computed for the next period.

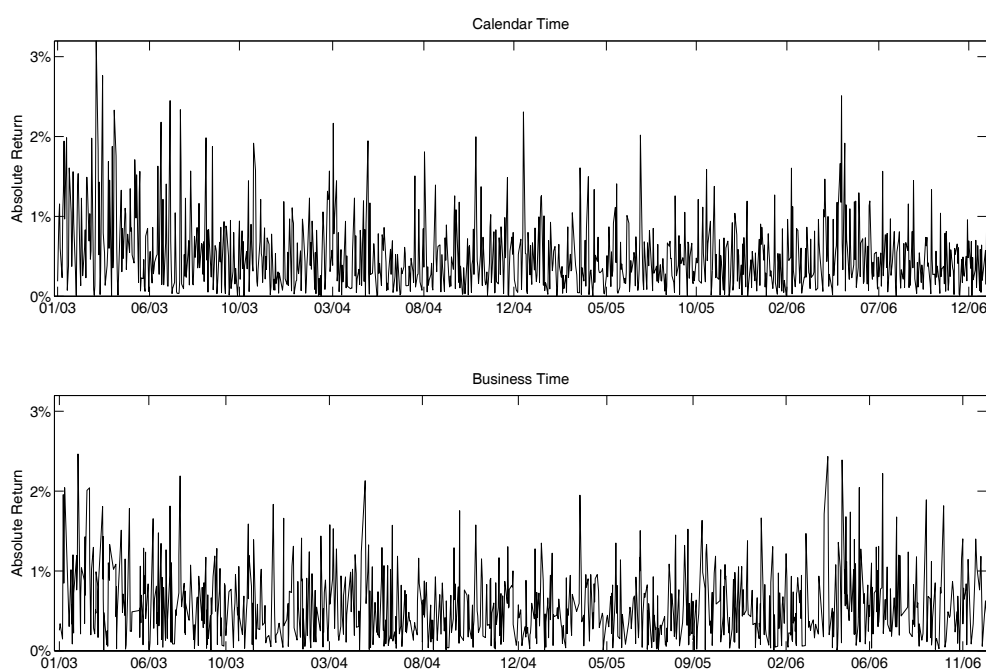
The results are rather mixed: both calendar-time and business-time minimum portfolios yield an annualized mean absolute returns of about 8.5% (see Figures 4.15 to 4.16). Out-of-sample returns, however, shows that passing in business time decreases the tails of the distribution. This result remains a consequence of the translation of the portfolio to business time and does not seem to reflect a significantly better measure of the underlying correlations.

Figure 4.14. Clustering of Stocks in Calendar and Business Times



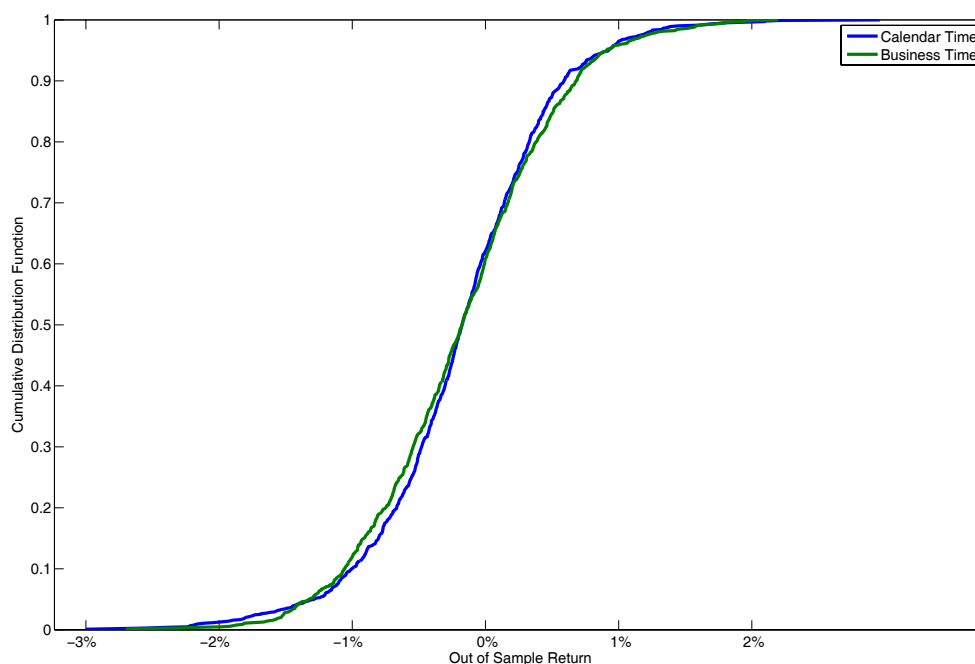
Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. The figure represents the level of hierarchical classification of stocks in calendar and business times.

Figure 4.15. Out-of-sample Absolute Returns of the Minimum Variance Portfolios



Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. Out-of-sample returns of the Minimum Variance Portfolio computed with 260 rolling observations.

Figure 4.16. Distribution Function of Out-of-Sample Returns of the Minimum Variance Portfolios



Source: Euronext, 5' intraday 43 stocks sample quotes on the period 01/02/2002 to 12/29/2006. Out-of-sample returns of the Minimum Variance Portfolio computed with 260 rolling observations.

4.5 Conclusion

The aim of this chapter is twofold: first to present a practical scheme to retrieve an implicit business time corresponding to each financial time series, and second to show that the resulting transformation yields valuable results when used in traditional financial applications.

So far, the first aim is achieved, the proposed genetic algorithm provides in every tested case a time-scale in which observed returns are Gaussian, according to a goodness-of-fit measure. Some issues nevertheless remain, like dependency on the choice of the exact measure or the non-uniqueness of the time-scale. These issues will be addressed in further developments by simultaneously using multiple *criteria* of normality, which shall

robustify the method and reduce the number of possible solutions.

An obvious application at the stock level is Value-at-Risk computation (see chapter 2), that could be simplified using the Gaussian approximation with arbitrary precision, we can think about for computing high-frequency VaR. Most applications, however, involve a combination of stocks to form a portfolio. In that case, we have shown that it is critical to define a common time scale for all variables. Otherwise the correlation structure is destroyed, both in returns and in durations. Unfortunately, the optimization scheme proves unpractical for a significant number of stocks, and the resulting time-scale yields little improvements in terms of robustness of correlations, though it still produces thinner tails. Developments for the multivariate model shall therefore involve fewer stocks, possibly picked according to a prior analysis of their correlations, so that the common time-scale is not forced to accommodate radically different information clocks.

Conclusion Générale

Au travers de quatre essais empiriques, nous avons dans cette thèse proposé des méthodes adaptées à l'exploitation efficiente des données de haute fréquence dans les principaux domaines de la littérature financière (microstructure, gestion de portefeuilles, mesure de risques et distribution des rendements).

Ces données de haute fréquence présentent de nombreuses particularités qu'il convenait de caractériser avant d'appliquer les méthodes de la finance quantitative. Nous avons ainsi mis en évidence un phénomène microstructurel (la « rose des vents ») sur les séries en haute fréquence de quelques titres du marché français des actions. La cause naturelle de ce type de phénomène repose, comme nous l'avons montré, sur la discrétisation des cours de bourses due aux frictions du processus de fixation des prix. Plusieurs auteurs (Krämer et Runde, 1997 et Amilon et Byström, 2000) ont montré les conséquences de la discrétisation des prix sur les estimations des modèles de la théorie financière ; le phénomène de rose des vents peut, à partir d'un seuil (non déterminé explicitement jusqu'ici), être la cause de biais sur les estimations des modèles autorégressifs du type *ARCH* et sur les tests standards de chaos.

Nous avons proposé une caractérisation de ce phénomène à l'aide de trois mesures quantitatives complémentaires originales et nous avons montré par simulations de processus bien connus des rendements, que cette structure apparaissait systématiquement à partir d'un seuil de pas d'échelon de cotation élevé. Cette approche d'analyse descrip-

tive des données de haute fréquence par l'étude d'un phénomène de microstructure - non encore étudié sur le marché français - nous a permis de proposer une mesure et d'illustrer son impact sur l'estimation des modèles financiers. Nous avons ainsi pu définir les plages de fréquence d'observations, qui, tout en limitant les effets des bruits dus à des phénomènes de microstructure, maximisent le contenu informationnel des cours boursiers.

La première application directe que nous avons réalisé sur les données de haute fréquence porte sur les mesures de risque, avec pour objectif d'affiner les mesures traditionnelles, à l'aide des données plus précises et des méthodes de calcul adaptées. La mesure de risque utilisée le plus souvent dans la littérature est la Valeur-en-risque (*Value-at-Risk*, *VaR*). Celle-ci suppose la connaissance de la densité de probabilité des rentabilités anticipées de l'actif étudié, qui peut s'approximer, entre autres, par des techniques de génération de *scenarii*. Généralement l'Analyse en Composantes Principales est utilisée pour extraire le nombre de *scenarii* des données. Néanmoins, le filtre linéaire de l'Analyse en Composantes Principales conduit à une perte d'information lorsque les données sont mesurées en haute fréquence. Nous avons proposé ici une mesure complémentaire définie à partir de l'Analyse en Composantes Indépendantes qui est appropriée lorsque la non-linéarité et la non-normalité sont inhérentes aux données (*Cf.* Back et Weigend, 1997). En utilisant des données de haute fréquence de quelques titres vifs sur le marché français des actions, nous trouvons que l'hypothèse de l'existence des facteurs de risques globaux caractérisant le marché est bien vérifiée. De plus, les mesures de risque calculées à partir des facteurs issus d'une Analyse en Composantes Indépendantes (*ICA-VaR*) sont plus robustes que celles déduites des séries originales ou bien de l'ACP.

L'application suivante portait sur les modèles d'évaluation d'actifs financiers. Les modèles classiques reposent, dans la majorité des cas, sur l'hypothèse de « quasi-normalité » ; mais cette hypothèse ne correspond pas aux faits stylisés. En effet, les variables financières sont caractérisées par de fortes particularités : queues de distributions épaisses, asymétrie de la distribution et hétéroscédasticité. Les applications basées sur les données de haute

fréquence peuvent, dans ce contexte et notamment en présence de variations des paramètres, être utilisées pour l'estimation des paramètres de basse fréquence des modèles financiers et pour l'allocation d'actifs. Comme nous l'avons montré, les données de haute fréquence contiennent du bruit, il fallait donc nécessairement développer des modèles appropriés pour pleinement exploiter leur potentiel.

Dans ce sens, nous avons utilisé des estimations de mesures de sensibilité effectuées sur le marché français des actions et ensuite, suivant l'approche développée récemment (Bollerslev et Zhang, 2003), nous avons estimé plusieurs modèles classiques d'évaluation des actifs notamment le modèle d'équilibre des actifs financiers ainsi que des extensions de leurs versions originales. Ces extensions comprennent principalement les modèles d'évaluation des moments d'ordre supérieur et les modèles conditionnels d'évaluation d'actifs. Les estimations ont été menées avec des méthodologies qui permettent de neutraliser les erreurs de spécification et de mesure des données tout en tenant compte des inter-relations entre les actifs financiers en terme de rentabilité, mais aussi de moments conditionnels d'ordre supérieur. Nous avons ainsi montré que les résultats sont sensiblement améliorés par rapport à une approche classique avec des estimations à partir des données de basse fréquence.

Nous avons finalement proposé un algorithme permettant de ré-échantillonner les données à haute fréquence de manière à obtenir des séries ayant des caractéristiques distributionnelles gaussiennes. Les nouvelles séries comprennent des observations espacées irrégulièrement dans le temps, mais régulières dans un temps déformé, appelé temps d'échange (ou de transaction) implicite, dans lequel la distribution des rendements sera proche de la loi normale. Comme rappelé précédemment, les caractéristiques des rendements calculés en temps calendaire sont connues pour être incompatibles avec les hypothèses de Gauss-Markov, excepté à de très basses fréquences (moins d'une observation par semaine). Une première manière de résoudre ce problème est d'échantillonner les prix à différentes dates suivant des variables exogènes. Le temps d'échange est alors déduit

des variables extraites à partir des séries de rendements, telle que la volatilité instantanée ou de certaines variables complémentaires.

Nous décrivons alors un procédé pour extraire ce temps d'échange sur des séries temporelles des prix à partir d'une méthode de recherche directe (algorithme génétique). Nous vérifions que l'algorithme proposé permet, dans tous les cas, d'obtenir des rentabilités gaussiennes à des fréquences où l'hypothèse de normalité est rejetée sans ambiguïté pour l'échantillonnage calendaire. Les applications pratiques de cet algorithme tirent partie de la normalité des séries obtenues, ce qui permet alors d'appliquer directement les formules issues de la finance traditionnelle après complétion de l'algorithme. Les rendements les plus proches de la distribution gaussienne sont utilisés pour le calcul de la matrice de variance-covariance et l'allocation des portefeuilles, et comparés avec les portefeuilles issus des temps calendaires. Les résultats donnent des variations parfois substantielles dans les corrélations et variances mesurées. L'étude des performances de ces allocations en dehors de l'échantillon ne montre néanmoins pas de surperformances significatives. Cependant, l'exploration des résultats montre l'importance, dans la définition du temps d'échange, d'un regroupement préliminaire d'actifs de sensibilités proches, de manière à éviter à la fois des désynchronisations trop fortes entre les titres, et une normalisation forcée des actifs hétérogènes.

L'ensemble des travaux conduits dans le cadre de cette thèse montre la valeur ajoutée, pour des applications financières, de l'utilisation de données de haute fréquence. L'enjeu véritable, lié à l'utilisation systématique de ces données, est le gain de précision que celles-ci permettent d'obtenir dans les applications empiriques. Cette prise en compte des caractéristiques intrinsèques des rendements des actifs en haute fréquence (biais microstructuraux, non-linéarités, distributions asymétriques, queues de distribution épaisses, ...) se traduit, comme nous l'avons montré, par une meilleure stabilité et précision des alloca-

tions et des mesures de risques.

Enfin, plusieurs pistes d'approfondissement et d'amélioration sont en cours d'études. En effet, nous comptons étendre l'étude de la rose des vents sur l'ensemble des places boursières composant l'*Euronext* (Amsterdam, Bruxelles, Lisbonne et Paris). Nous disposons désormais d'une base de données de transactions provenant d'*Euronext*, allant du 1^{er} janvier 2002 au 31 décembre 2007 contenant tous les titres. Nous pourrions ainsi tester la présence du phénomène sur les autres marchés en tenant compte des mouvements récents des prix des actifs. Les principales valeurs limites des indicateurs de roses des vents, déterminées sur quelques titres du marché français au premier semestre 2004, pourront ainsi être mises à jour et généralisées à l'échelle européenne. Nous envisageons aussi de faire une analyse comparative des différentes mesures de visibilité du phénomène de rose des vents disponibles dans la littérature (*Cf.* Koppl et Nardone, 2001 ; Wang et Wang, 2002 et Mitchell et McKenzie, 2006). Il serait aussi intéressant d'étudier le comportement des paramètres des modèles *GARCH* de manière empirique par simulations sur des données continues (sans pas d'échelon de cotation) et sur des données discrétisées avec plusieurs pas d'échelon de cotation différents. Nous pourrions alors à partir des indicateurs de rose des vents, déterminer l'ampleur des erreurs d'estimation engendrées par la discrétisation et la structure de corrélation imposée. Nous pourrions pour cela exploiter les travaux de Amilon et Byström (2000) et de Amilon (2003), qui montrent comment le phénomène biaise les principaux tests statistiques.

De même, les résultats obtenus sur le calcul de la *VaR* en haute fréquence montrent bien que cette technique est une alternative à l'Analyse en Composantes Principales dans les cas où l'hypothèse de distribution gaussienne des rendements est approximative. Quelques améliorations et extensions peuvent être apportées. La première extension provient de la modélisation des facteurs de risque déterminés par l'algorithme. Dans des travaux en cours, nous testons des familles de distributions intégrant davantage les moments d'ordre supérieur ; notamment à partir des modèles *Skew-student* (Keith et Shen,

2004 et Adcock et Meade, 2007). Une autre piste de travail, est de tester d'autres approches d'extraction des facteurs indépendants comme l'utilisation des copules (*COPICA* pour *COPula Independent Component Analysis* en anglais, Cf. Chen *et alii* 2008) ou bien les méthodes avec contraintes. En effet, la version de l'ACI présentée dans ce chapitre ne permet pas, comme dans le cas d'une ACP, de réduire le nombre de facteurs ; elle détermine automatiquement le nombre de facteurs statistiquement indépendants. Nous pouvons à partir d'une Analyse en Composantes Indépendantes avec contraintes (*cICA* pour *constraint Independent Component Analysis* en anglais, Cf. Lu et Rajapakse, 2000 et 2005), réduire la dimension des données tout en gardant le caractère d'indépendance des facteurs. De plus, nous pouvons supposer que les données observées sont des combinaisons des sources indépendantes qui sont affectées par des bruits. A cet effet, Chang et Zhang (2006) et Kopriva et Seršić (2008) ont proposé des algorithmes qui permettent de filtrer les bruits par la méthode des ondelettes (*Wavelet* en anglais) avant de déterminer les composantes indépendantes. Cette forme d'extraction de facteurs indépendants (*SDICA* pour *Subband Decomposition Independent Component Analysis* et *WP-SDICA* pour *Wavelet Packets SDICA* en anglais) fournit directement la bande dé-bruitée de fréquences dans laquelle les sous-composantes du signal d'origine sont indépendantes. Cette dernière technique applique un filtre aux observations pour permettre de décomposer l'échantillon en bandes de sous-composantes indépendantes et dépendantes.

Il serait ensuite intéressant d'évaluer formellement l'importance des erreurs de mesure des sensibilités aux facteurs dans le contexte actuel de mesures et co-mesures réalisées pour le *CAPM*. Bai *et alii* (2000), Barndorff-Nielson et Shephard (2002a, b), Meddahi (2002) ou récemment Griffin et Oomen (2008), fournissent des résultats théoriques pour des mesures réalisées de volatilité. La volatilité instantanée est alors décomposée en une volatilité réalisée et terme représentant le biais d'auto-corrélation. Griffin et Oomen (2008) montrent théoriquement et empiriquement que le biais d'auto-corrélation diminue avec la fréquence d'échantillonnage des prix pour être négligeable après 25 minutes. Un travail similaire pourrait donc être fait sur le *beta* réalisé et la *skewness* réalisée. De

même, en utilisant un modèle à facteurs indépendants, les différentes relations entre les actifs pourraient être simplifiées. En effet, l'extraction des facteurs communs entre les actifs par une Analyse en Composantes Indépendantes avec filtrage de bruits (*Cf.* Kopriva et Seršić, 2008) engendrant des matrices des cumulants diagonales (du fait de la propriété d'indépendance des facteurs). De ce fait, l'estimation des paramètres avec les moments d'ordre supérieur pourra s'effectuer de manière simple avec les méthodes traditionnelles. Enfin, inspirée par les travaux de Brock et Hommes (1998) ou de Malevergne et Sornette (2006), la généralisation des approches précédentes dans un cadre hétérogène - où les agents confrontent leur fourchette de prix - pourrait constituer une extension naturelle de cette étude. Pour cela, nous pouvons envisager plusieurs sources d'hétérogénéité. Les agents peuvent alors avoir des points de vue différents par rapport aux anticipations des prix (Brock et Hommes, 1998). Nous pouvons aussi supposer que l'hétérogénéité provient de la perception du risque par les agents. De ce fait, plusieurs mesures de risque pour les allocations optimales entre actifs risqués et non risqués peuvent être utilisées conjointement dans un cadre multi-varié (Malevergne et Sornette, 2006).

Finalement, les futures études sur l'évaluation d'actifs, l'attribution de performances ou bien la gestion des risques pourraient s'appuyer sur des données implicites. Dans ce cas de figure, l'invariance recherchée dans la distribution des rendements est le résultat d'une compréhension globale des fluctuations des prix et des fluctuations du temps. Le temps boursier n'est plus alors indépendant de l'action des opérateurs, mais étroitement lié aux variables comme la taille du marché, le volume de l'échange et le nombre de participants sur le marché. L'activité des opérateurs agit alors comme une pondération des durées, en contractant plus ou moins l'espace des cotations. Le temps calendaire physique exogène du marché devient un temps intrinsèque endogène au marché. Il est possible d'envisager un temps implicite pour chacune de ces variables et une agrégation multi-variée de ces différents temps pour une meilleure prise en compte des vraies sources de variations des prix des actifs. De ce fait, la valeur d'un prix coté (et donc la pertinence de la mesure) n'est pas la même selon que ce prix correspond à une transaction de cinq cent mille titres

ou de cinq titres. Cette intuition qui trouve une traduction scientifique dans les modèles à temps déformé traduit le fait que sur les marchés financiers, la durée du temps n'est pas égale à la valeur du temps. Toutefois, le temps implicite déterminé par les algorithmes proposés dans ce chapitre ne fait que transférer le problème de l'hypothèse distributionnelle en temps calendaire au problème de prévision des prix et correspondance des valeurs en temps déformé. Néanmoins, tout comme pour les modèles à duration (Engle et Russell, 1997), où les informations sur les prix ne sont pas échantillonnées sur un intervalle de temps régulier, le temps implicite pourrait servir de base à de nouveaux modèles sur un espace particulier (espace du volume des échanges, du nombre de participants ou bien du niveaux de variation du prix). De plus, de nouveaux produits financiers pour des investisseurs particuliers, exprimés en temps implicite, pourraient voir le jour. Ces produits pourront alors être définis suivant des critères exprimés en temps déformé (volatilité constante sur cet espace, valeurs liquidatives enregistrées suivant le volume atteint, ...).

Annexe :

On the Impact of the Compass Rose Phenomenon on GARCH Estimation

1. Introduction

The Scatter Plot of stock returns against their lagged values reveals a striking pattern of lines radiating from the origin and pointing in the major directions of the compass. This phenomenon is the so-called “Compass Rose” of returns (Crack and Ledoit, 1996) and is the subject of many recent studies done with high frequency data (see Krämer and Runde, 1997; Szpiro, 1998 and Lee *et al.*; 1999, 2005), aiming to discover its causes and determinants. The structural cause of the Compass Rose is found into the discretization of the asset price. Crack and Ledoit (1996) show that the tick size is crucial for this phenomenon on the stocks traded on the New York Stock Exchange (NYSE). Amilon and Byström (2000), Ball (1988), Amilon (2003) and Krämer and Runde (1997) show, in the subjective analysis, that discretization has some effect on the estimation of parameters when using high frequency. The purpose of this work is to test with a quantitative measure and simulated data the effect of Compass Rose structure in the GARCH models estimation and to find the optimal sub-sampling frequency for the parameters estimations in the high frequency context. We firstly give the simulation framework (Section 2) and present after, the experiment and results (Section 3); section 4 concludes.

2. The Models

In this section, we present the framework of our simulation studies on a General AutoRegressive Conditional Heteroskedasticity type model (Bollerslev, 1986) simulation and the Compass Rose Intensity Indexes.

2.1 The GARCH Simulation

We perform simulation experiments under the hypothesis that the high frequency volatility is serially uncorrelated but heteroskedastic, for evaluating the compass rose effect on parameters estimations. Let the P_t be the unobserved equilibrium price of a stock at time t and r_t the unobserved return of the stock. The continuous-state *GARCH*(1,1) model takes the following form:

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta \sigma_{t-1}^2$$

where parameters $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta > 0$ and σ_t^2 is the conditional variance of the return r_t .

We compare the parameters values of the estimation of the continuous GARCH model to handle case where the return series are calculated from discrete prices. The discretization of the continuous price is performing with the following methodology:

$$P_t^* = Ent\left(\frac{P_t}{d} + 0.5\right) d$$

where $Ent(.)$ is the floor function, the parameter d is the tick size and P_t^* is the corresponding discrete price of the continuous price P_t .

2.2 The Compass Rose Indexes

For measuring the Compass Rose Phenomenon, we use some definition of a Compass Rose Index¹.

¹See Appendix 1 for formal definitions of the measures.

The main idea behind the first proposed index is that in the case of a perfect Compass Rose all returns lie on four and only four lines representing the quadrant directions: North-South, North East-South West, North West-South East and East-West. Each observation of a pair of returns in the scatter plot, together with the origin, defines a line. This last Index ($CRII_1$) is then linked to the proportion of lines originating from 0 and being on the four perfect directions mentioned above.

The second quantitative measure (denoted $CRII_2$) accordingly computes the concentration of angles corresponding to the main lines of the Compass rose, compared to the distribution of possible angles thus giving the distortion caused by the discretization effect (Koppl and Nardone, 2001).

The third Compass Rose Intensity Index ($CRII_3$) proposed is based on the fact that this phenomenon capture a discretization effect and a correlation effect, but depends upon a traditional contingency table. This Index is computed directly from the contingency table. When a Compass Rose is present, the density function of return will be different from the theoretical uniform distribution and the density function of the bootstrapped series of returns.

3. Simulations and Estimations

In order to test the effect of the temporal heterogeneity of the market, we use the maximum likelihood procedure to fit the simple $GARCH(1,1)$ model (see equation 1) for several tick size. We first simulate 30,000 continuous observations with the parameters $(2.2 \times 10^{-5}; 0.11; 0.83)$. Parameters obtained by Amilon and Byström (2000) for “Atlas copco A fria” stock in the Stockholm market. This prices series are used to evaluate the optimal sub-sampling and the impact of the discretization.

3.1 The Optimal Sub-sampling

To evaluate the existence of optimal sub-sampling, we use tick size between 1 and 5 to rounding the data following equation 2 procedure. Figure A.1 shows the simulated prices (column 1) and the corresponding returns (column 2).

- Please insert Figure A.1 somewhere here -

The Compass Rose Intensity Indexes are used to evaluate the discretization level of the previous three series in the various sub-sampling schemes. As could be expected, the indexes have high value for the high tick size levels (see Figure A.2). We also see that the indexes of the continuous and the rounded series converge to the same value when the sub-sampling increases. This result means that the optimal sub-sampling of the rounding data depends on the tick size (Crack and Ledoit, 1996; Krämer and Runde, 1997; Amilon and Byström, 2000).

- Please insert Figure A.2 somewhere here -

3.2 The Rounding Impact on GARCH Estimation

The continuous time model may provide a better approximation to the discrete data than a discrete model. In this case we can find a sequence of processes, whose increments on finer becoming grids are GARCH processes (Nelson, 1990; Nelson and Foster, 1994). In both cases, the sum of the α (ARCH effect) and β (GARCH effect) should tend to 1 as the frequency becomes larger. Drost and Nijman (1993) confirm the adequacy of these theoretical results at the daily and weekly frequencies and Guillaume *et al.* (1995) with intra-day data. But, when the prices are limited to move in the some restricted space, can we obtain the same sequence of processes? We use in this paper a high frequency simulated data for testing the stability of the coefficient estimates for several tick size level (see Figure A.3).

- Please insert Figure A.3 somewhere here -

As can be seen, the use of discrete data for GARCH model cause a complete break-down of the estimation procedure when the tick size increases, since the coefficients are not stable over the tick size. Moreover, the sum of the parameters does not really tend to 1 for the tick size 0.5 to 4; and the persistence of the volatility and maximum likelihood decreases on tick size between 0.5 and 1 (see Figure A.4). This means that, in the presence of discrete data, the GARCH model does not have the same parameters with the continuous one. But, when the tick size is reasonably small compared to the price level, the differences are not serious (on the contrary, with higher tick size, the model is unstable).

- Please insert Figure A.4 somewhere here -

4. Conclusion

The main purpose of this paper is to provide a comprehensive analysis of prices discretization in the time series analysis. We use simulated data to show how the tick size can modify the structure of parameters described by a GARCH model. In this context, we see that the relatively small tick size does not modify significantly the parameters, but in other way, the model become unstable. This result obtained with Monte Carlo simulations confirms the conclusion of Krämer and Runde (1997), Amilon and Byström (2000), Koppl and Tuluca (2004) and Amilon (2003) regarding the bias in estimation. We also show that for all tick size level, we obtain some sub-sampling that gives small value of quantitative measures of discretizaion effect. In this case, the corresponding value of sub-sampling can be view like optimal for estimations of statistical models in high frequency context or in low frequency with high frequency data.

Appendix 1: The Compass Rose Indexes

The first Index ($CRII_1$) is linked to the proportion of lines originating from 0 and being on the four perfect directions representing the quadrant directions: North-South, North East-South West, North West-South East and East-West. Formally:

$$CRII_1 = 4/\{p [\#(D_0) - 4] + 4\} \quad (4.1)$$

where $\#(D_0)$ is the number of all the lines passing by the origin and p (with $0 < p < 1$) an arbitrary constant.

The second quantitative measure is Compass Rose Intensity 2 (denoted $CRII_2$), defined as follows:

$$CRII_2 = \sum_{w=-1}^1 \left\{ \frac{[\hat{f}(w) - u(w)]^2}{u(w)} \right\} + \sum_{w=-1}^1 \left\{ \frac{[\hat{f}(w) - \hat{g}(w)]^2}{\hat{g}(w)} \right\} \quad (4.2)$$

where $\hat{f}(\cdot)$ is the empirical density of polar coordinates w , $u(\cdot)$ is the uniform distribution and $\hat{g}(\cdot)$ the empirical density function of the bootstrap of polar coordinates w , and:

$$\omega_t = \theta_t/\pi = \begin{cases} [\arctan(r_t/r_{t-1})]/\pi & \text{if } (r_t \geq 0) \\ [\arctan(r_t/r_{t-1}) + \pi]/\pi & \text{if } (r_t < 0) \text{ and } (r_t \geq 0) \\ [\arctan(r_t/r_{t-1}) - \pi]/\pi & \text{if } (r_t < 0) \text{ and } (r_t < 0) \end{cases}$$

with $-\frac{\pi}{2} \leq \arctan(x) \leq \frac{\pi}{2}$ for all $x \in \mathbb{R}$.

The third Compass Rose Intensity Index ($CRII_3$) is written as:

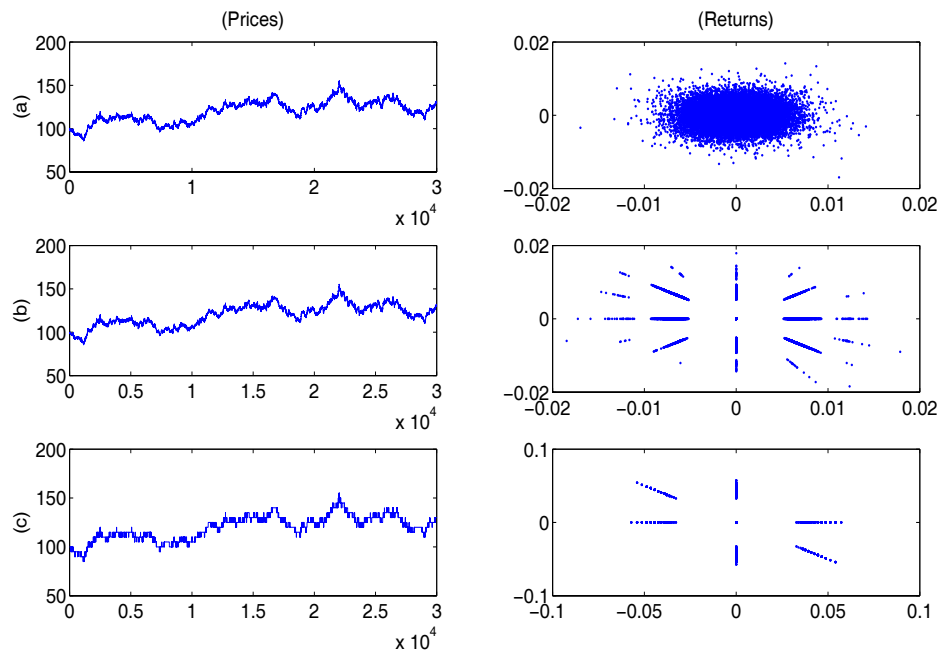
$$CRII_3 = \|z\|^2 = \sum_{i=1}^k \sum_{j=1}^k (\hat{p}_{ij} - u_{ij})^2 + \sum_{i=1}^k \sum_{j=1}^k (\hat{p}_{ij} - p_{ij})^2 \quad (4.3)$$

with \hat{p}_{ij} the density function of the return in the cells (i, j) of the contingency table, u_{ij}

the theoretical uniform distribution, p_{ij} the density function of bootstrapped series of the original return and some optimal partition of the contingency table.

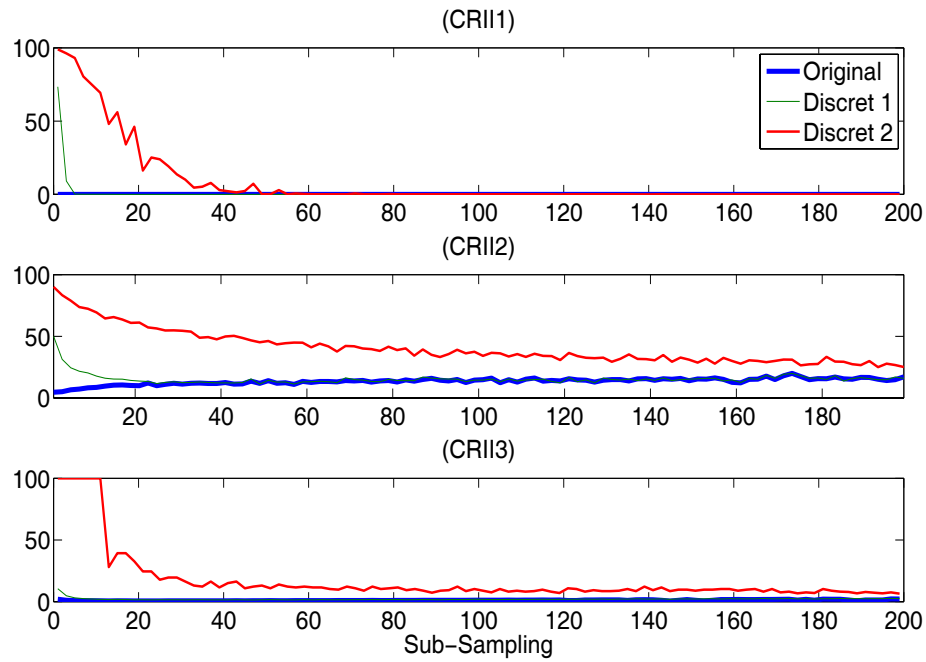
Appendix 2: List of Figures

Figure A.1. Prices and Corresponding Returns of the Simulated data



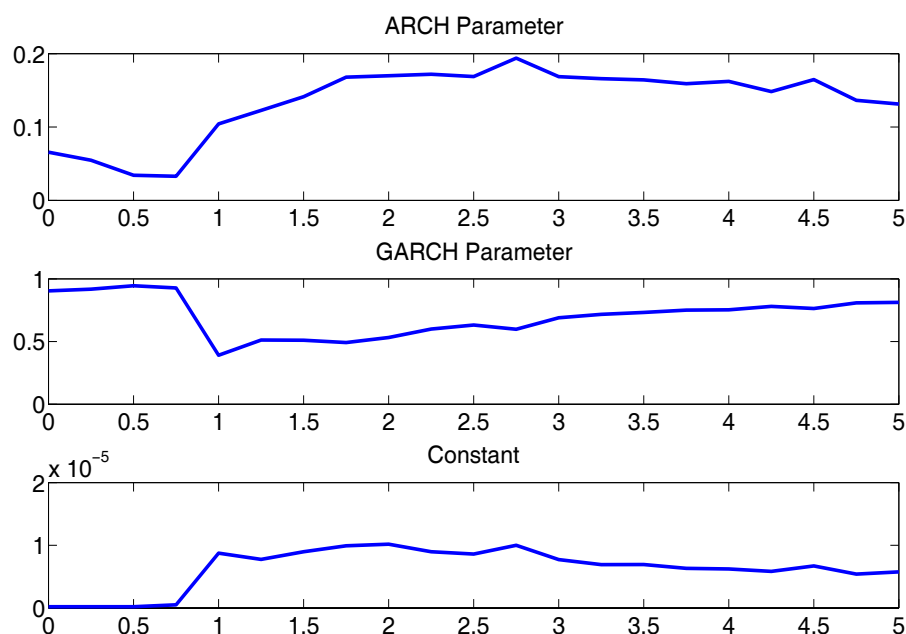
Source: simulation of 30,000 observations with $GARCH(1,1)$ model; the parameters are $\{2.2 \times 10^{-5}; 0.11; 0.83\}$ (Parameters obtained by Amilon and Byström, 2000, for “Atlas copco A fria” stock in the Stockholm market). This figure shows the simulated prices (column 1) and the corresponding returns (column 2) with the tick size 0 - original prices - line (a); the tick size 1, line (b) and the tick size 5, line (c).

Figure A.2. Evolution of the Compass Rose Intensity Indexes under Sub-sampling times



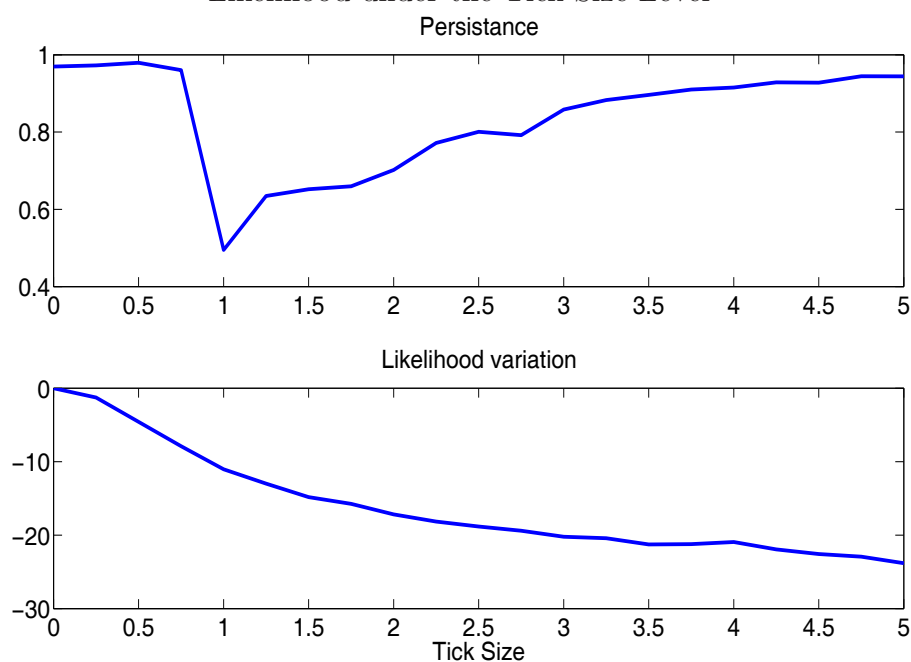
Source: simulation of 30,000 observations with $GARCH(1,1)$ model; the parameters are $\{2.2 \times 10^{-5}; 0.11; 0.83\}$ (Parameters obtained by Amilon and Byström, 2000, for “Atlas copco A fria” stock in the Stockholm market). This figure shows the Compass Rose Intensity Index 1 (top), the Compass Rose Intensity Index 2 (middle) and the Compass Rose Intensity Index 3 (bottom) under the level of sub-sampling. The blue line is the computation on the original simulated prices, the green line is the rounding prices with the tick size 1 and the red line is the rounding prices with the tick size 5.

Figure A.3. Evolution of the GARCH Parameters Estimation under the Tick Size Level



Source: simulation of 30,000 observations with $GARCH(1,1)$ model ; the parameters are $\{2.2 \times 10^{-5}; 0.11; 0.83\}$ (Parameters obtained by Amilon and Byström, 2000, for “Atlas copco A fria” stock in the Stockholm market). This figure shows the ARCH parameters estimation (top), the GARCH parameters estimation (middle) and the constant of the model (bottom) under the tick size (0 to 5).

Figure A.4. Evolution of the Persistence of the GARCH Model and Variation of the Likelihood under the Tick Size Level



Source: simulation of 30,000 observations with $GARCH(1,1)$ model ; the parameters are $\{2.2 \times 10^{-5}; 0.11; 0.83\}$ (Parameters obtained by Amilon and Byström, 2000, for “Atlas copco A fria” stock in the Stockholm market). This figure shows the persistence of the GARCH model - sum of all the parameters - (top) and the likelihood variation of the model (bottom) under the tick size (0 to 5).

Références bibliographiques

Aït-Sahalia Y., P. Mykland et L. Zhang, (2005), “How often to Sample a Continuous-time Process in the Presence of Market Microstructure Noise”, *Review of Financial Studies* 18, 351-416.

Abramowitz M. et I. Stegun, (1964), *Handbook of Mathematical Functions*, Dover, 1408 pages.

Aitken M. et A. Frino, (1996), “Asymmetry in Stock Returns following Block Trades on the Australian Stock Exchange: A Note”, *Abacus* 32, 54-61.

Albanese C., K. Jackson et P. Wiberg, (2002), “Dimension Reduction in the Computation of Value-at-Risk”, *Journal of Risk Finance*, 1-13.

Amari S., S. Cichocki et A. Young, (1996), “A New Learning Algorithm for Blind Signal Separation”, in *Advances in Natural Information Processing Systems* 8, Touretzky-Mozer-Hasselmo Eds, MIT Press, Cambridge MA, 757-763.

Amilon H., (2003), “GARCH Estimation and Discrete Stock Prices: An Application to Low-priced Australian Stocks”, *Economics Letters* 81, 215-222.

Amilon H. et H. Byström, (2000), “The Compass Rose Pattern of the Stock Market: How Does it Affect Parameter Estimates, Forecasts, and Statistical Tests?”, *Working paper*, Lund University, 22 pages.

Ané T. et H. Geman, (2000), “Order Flow, Transaction Clock and Normality of Asset Returns”, *Journal of Finance* 55, 2259-2284.

Andersen T., T. Bollerslev et A. Das, (2001), “Variance-ratio Statistics and High-frequency Data: Testing for Changes in Intraday Volatility Patterns”, *Journal of Finance* 56(1), 305-327.

Andersen T., T. Bollerslev et F. Diebold, (2002), “Parametric and Nonparametric Volatility Measurements”, in *Handbook of Financial Econometrics*, Hansen-Aït Sahalia Eds, North-Holland Amsterdam, 68 pages.

Andersen T., T. Bollerslev, F. Diebold et P. Ebens, (2001), “The Distribution of Realized Stock Return Volatility”, *Journal of Financial Economics* 61, 43-76.

- Andersen T., T. Bollerslev, F. Diebold et P. Labys, (2000), "Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian", *Multinational Finance Journal* 4, 159-179.
- Andersen T., T. Bollerslev, F. Diebold et P. Labys, (2002), "Modeling and Forecasting Realized Volatility", *Econometrica* 71(2), 579-625.
- Andersen T., T. Bollerslev et S. Lange, (1999), "Forecasting Financial Market Volatility: Sample Frequency *vis-a-vis* Forecast Horizon", *Journal of Empirical Finance* 6, 457-477.
- Annaert J., A. Claes et M. de Ceuster, (2003), "Does the Compass Rose Pattern Matter for Testing Normality?", *Working paper*, Deloitte & Touche, 14 pages.
- Anshuman V. et A. Kalay, (1998), "Market Making with Discret Prices", *Review of Financial Studies* 11(1), 81-109.
- Antoniou A. et C. Vorlow, (2005), "Price Clustering and Discreteness: Is there Chaos behind the Noise?", *Physica A: Statistical Mechanics and Its Applications* 348, 389-403.
- Artzner P., F. Delbaen, J. Eber et D. Heath, (1998), "Coherent Measures of Risk", *Finance and Stochastics* 9, 203-228.
- Avouyi-Dovi S., D. Guégan et S. Ladoucette, (2002), "Une mesure de la persistance dans les indices boursiers", *Notes d'études et de recherche* 94, Banque de France, 46 pages.
- Back A. et A. Weigend, (1997), "A First Application of Independent Component Analysis to Extracting Structure from Stock Returns", *International Journal of Neural Systems* 8, 1-16.
- Back A. et T. Trappenberg, (1999), "Input Variable Selection using Independent Component Analysis", *International Joint Conference on Neural Networks*, Washington, 4 pages.
- Bai X., J. Russell et G. Tiao, (2000), "Beyond Merton's Utopia: Effects of Non-Normality and Dependence on the Precision of Variance Estimates Using High-frequency Financial Data", *Working paper*, Graduate School of Business, University of Chicago, 48 pages.
- Bali T. et L. Wu, (2005), "Estimating the Intertemporal Capital Asset Pricing Model with Cross-sectional Consistency", *Social Science Research Network*, 56 pages.
- Ball C., (1988), "Estimation Bias induced by Discrete Security Prices", *Journal of Finance* 43, 841-865.
- Banz R., (1981), "The Relationship between Return and Market Value of Common Stocks", *Journal of Financial Economics* 9, 3-18.
- Barndorff-Nielsen O. et N. Shephard, (2002a), "Econometric Analysis of Realized Covariation: High Frequency Covariance, Regression and Correlation in Financial Economics", *Econometrica*

72(3), 885-925.

- Barndorff-Nielsen O. et N. Shephard, (2002b), "Estimating Quadratic Variation using Realized Variance", *Journal of Applied Econometrics* 17, 457- 477.
- Barone A. et P. Talwar, (1983), "Market Models and Heteroskedasticity of Residual Security Returns", *Journal of Business and Economic Statistics* 1, 163-168.
- Barone A., (1985), "Arbitrage Equilibrium with Skewed Asset Returns", *Journal of Financial and Quantitative Analysis* 20, 299-313.
- Barone A., K. Giannopoulos et L. Vosper, (2002), "Backtesting Derivative Portfolios with Filtered Historical Simulation (FHS)", *European Financial Management*, 31-58.
- Barone A., P. Gagliardini et G. Urga, (2003), "Testing Asset Pricing Models with Coskewness", *Journal of Business and Economic Statistics* 22(4), 474-485.
- Beine M, S. Laurent et F. Palm, (2004), "Central Bank Forex Interventions Assessed using Realized Moments", *Working Paper*, CORE, Université Catholique de Louvain, 39 pages.
- Bell A. et T. Sejnowski, (1995), "An Information Maximization Approach to Blind Separation and Blind Deconvolution", *Neural Computation* 7, 1129-1159.
- Berkowitz J. et F. Diebold, (1998), "Bootstrapping Multivariate Spectra", *Review of Economics and Statistics* 80, 664-666.
- Berkowitz J., (2001), "Testing Density Forecasts with Applications to Risk Management", *Journal of Business and Economics Statistics* 19, 465-474.
- Biais B., (1993), "Price Formation and Equilibrium Liquidity in Fragmented and Centralized Markets", *Journal of Finance* 38(4), 1053-1074.
- Black F., (1972), "Capital Market Equilibrium with Restricted Borrowing", *Journal of Business* 45, 444-55.
- Black F., M. Jensen et M. Scholes, (1972), "The Capital Asset Pricing Model: Some Empirical Tests", In *Studies in the Theory of Capital Markets*, Jensen Eds, Praeger Publishers, 79-121.
- Bollerslev T. et B. Zhang, (2003), "Measuring and Modeling Systematic Risk in Factor Pricing Models using High-frequency Data", *Journal of Empirical Finance* 10, 533-558.
- Bollerslev T., (1986), "Generalized AutoRegressive Conditional Heteroskedasticity", *Journal of Econometrics* 3, 307-327.
- Bollerslev T., R. Chou et K. Kroner, (1992), "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics* 52, 5-59.

- Bollerslev T., R. Engle et J. Wooldridge, (1988), "A Capital Asset Pricing Model with Time-varying Covariances", *Journal of Political Economy* 96, 116-131.
- Brandt M. et F. Diebold, (2004), "A No-arbitrage Approach to Range-based Estimation of Return Covariances and Correlations", *Working Paper 7*, Center for Financial Studies, 37 pages.
- Braun P., D. Nelson et A. Sunier, (1995), "Good News, Bad News, Volatility, and Betas", *Journal of Finance* 50, 1575-1603.
- Breeden D., (1979), "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities", *Journal of Financial Economics* 7, 265-96.
- Britten-Jones M. et S. Schaefer, (1999), "Non-linear Value-at-Risk", *European Finance Review* 2, 161-187.
- Brock W. et C. Hommes, (1998), "Heterogeneous Beliefs and Routes to Chaos in a Simple Asset Pricing Model", *Journal of Economic Dynamics and Control* 22, 1235-1274.
- Brock W., W. Dechert, J. Scheinkmann et B. LeBaron, (1996), "A Test for Independence based on the Correlation Dimension", *Econometric Reviews* 15, 197-235.
- Campbell J., A. Lo et A. MacKinlay, (1997), *The Econometrics of Financial Markets*, Princeton University Press, 622 pages.
- Campbell S., (2007), "A Review of Backtesting and Backtesting Procedures", *Journal of Risk* 9(2), 1-17.
- Capelle-Blancard G. et M. Chaudhury, (2007), "Price Clustering in the CAC40 Index Options Market", *Applied Financial Economics* 17(15), 1201-1210.
- Cardoso J. et A. Souloumiac, (1993), "Blind Beamforming for Non-gaussian Signals", *Proceeding of the IEE Conference* 140(6), 771-774.
- Cardoso J., (1989), "Source Separation using Higher Order Moments", *International Conference on Acoustics, Speech and Signal Processing*, 2109-2112.
- Cardoso J., (1990), "Eigen-structure of the Fourth-order Cumulant Tensor with Application to the Blind Source Separation Problem", In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2655-2658.
- Cardoso J., (1999), "High-order Contrasts for Independent Component Analysis", *Neural Computation* 11(1), 157-192.

- Cha S. et L. Chan, (2000), "Applying Independent Component Analysis to Factor Model in Finance", in *Intelligent Data Engineering and Automated Learning - IDEAL 2000, Data Mining, Financial Engineering, and Intelligent Agents*, Leung-Chan-Meng Eds, Springer, 538-544.
- Chan L. et J. Lakonishok, (1993), "Institutional Trades and Intraday Stock Price Behaviour", *Journal of Financial Economics* 33, 173-199.
- Chan L. et J. Lakonishok, (1995), "The Behaviour of Stock Prices around Institutional Trades", *Journal of Finance* 50, 1147-1174.
- Chan L. et J. Lakonishok, (1997), "Institutional Equity Trading Costs: NYSE *versus* NASDAQ", *Journal of Finance* 52, 713-735.
- Chen A., (1997), "The Square Compass Rose: The Evidence from Taiwan", *Journal of Multinational Financial Management* 7, 127-144.
- Chen R., M. Guo, W. Härdle et S. Huang, (2008), "Independent Component Analysis *via* Copula techniques", *Working Paper SFB 649*, Economic Risk, Berlin, 24 pages.
- Cheung Y. et L. Xu, (2001), "Independent Component Ordering in ICA Time-series Analysis", *Neurocomputing* 41, 145-152.
- Chidambaran N., C. Lee et J. Trigueros, (1999), "Option Pricing *via* Genetic Programming", *Computational Finance 1999*, Abu Mostafa-LeBaron-Lo-Weigend Eds, M.I.T Press, 583-598.
- Cho Y. et R. Engle, (1999), "Time-varying Betas and Asymmetric Effect of News: Empirical Analysis of Blue Chip Stocks", *Working Paper W7330*, National Bureau of Economic Research, 29 pages.
- Christoffersen P. et D. Pelletier, (2003), "Back-testing Value-at-Risk: A Duration-based Approach", *Journal of Financial Econometrics* 2(1), 84-108.
- Christoffersen P., (1998), "Evaluating Interval Forecasts", *International Economic Review* 39, 841-862.
- Clark P., (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica* 41(1), 135-155.
- Cochrane J., (2001), *Asset Pricing*, Princeton University Press, 530 pages.
- Colletaz G., C. Hurlin et S. Tokpavi, (2007), "Irregularly Spaced Intraday Value-at-Risk (ISIVaR) Models Forecasting and Predictive Abilities", *Working Paper*, University of Orléans, 37 pages.

- Comon P., (1994), "Independent Component Analysis - A New Concept?" *Signal Processing* 36(3), 287-314.
- Corsi F., (2006), "Realized Correlation Tick-by-Tick", *Working Paper*, University of Lugano, 21 pages.
- Corsi F., M. Dacorogna, U. Müller et G. Zumbach, (2001), "Consistent High-precision Volatility from High-frequency Data", *Economic Notes* 30, 183-204.
- Crack T. et O. Ledoit, (1996), "Robust Structure without Predictability: The Compass Rose Pattern of the Stock Market", *Journal of Finance* 51(2), 751-762.
- Dacorogna M., U. Müller, R. Gençay, R. Olsen et O. Pictet, (2001), *An Introduction to High-frequency Finance*, Academic Press, San Diego, CA, 383 pages.
- De Grauwe P. et D. Decupere, (1992), "Psychological Barriers in the Foreign Exchange Markets", *Journal of International and Comparative Economics* 1, 87-101.
- Demsetz H., (1968), "Why Regulate Utilities", *Journal of Law and Economics* 11(1), 55-65.
- Diebold F. et R. Mariano, (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics* 13, 253-264.
- Diebold F., T. Gunther et A. Tay, (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management", *International Economic Review* 39, 863-883.
- Dittmar R., (2002), "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from Cross Section of Equity Returns", *Journal of Finance* 57, 368-403.
- Drost F. et T. Nijman, (1993), "Temporal Aggregation of GARCH Processes", *Econometrica* 61(4), 909-927.
- Duffee D. et J. Pan, (1997), "An Overview of Value-at-Risk", *Journal of Derivatives* 4, 7-49.
- Duffee D. et J. Pan, (2001), "Analysis of Value-at-Risk with Jumps and Credit Risk", *Finance and Stochastics* 5, 155-180.
- Engle R. et J. Russell, (1997), "Forecasting the Frequency of Changes in Quoted Foreign Exchange Prices with the Autoregressive Conditional Duration Model", *Journal of Empirical Finance* 4, 187-212.
- Engle R. et S. Manganelli, (2004), "CAViaR: Conditional AutoRegressive Value-at-Risk by Regression Quantile", *Journal of Business and Economic Statistics* 22, 367-381.

- Epps T. et M. Epps, (1976), "The Stochastic Dependence of Security Price changes and Transaction Volume: Implication for the Mixture of Distributions Hypothesis", *Econometrica* 50, 305-321.
- Fama E. et J. MacBeth, (1973), "Risk, Return and Equilibrium: Empirical Tests", *Journal of Political Economy* 96, 607-636.
- Fama E. et K. French, (1993), "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics* 33, 3-56.
- Fama E. et K. French, (1995), "Size and Book to Market Factors in Earnings and Returns", *Journal of Finance* 50, 131-155.
- Fama E. et K. French, (1996), "Multifactor Explanations of Asset Pricing Anomalies", *Journal of Finance* 51, 55-84.
- Fang Y., (2002), "The Compass Rose and Random Walk Tests", *Computational Statistic and Data Analysis* 39, 299-310.
- Feinstone L., (1987), "Minute by Minute: Efficiency, Normality and Randomness in the Intradaily Asset Prices", *Journal of Applied Econometrics* 2, 193-214.
- Ferreira M. et J. Lopez, (2005), "Evaluating Interest Rate Covariance Models within a Value-at-Risk Framework", *Journal of Financial Econometrics* 3, 126-168.
- Ferson W. et C. Harvey, (1999), "Conditioning Variables and the Cross-section of Stock Returns", *Journal of Finance* 54, 1325-1360.
- Feuerverger A. et A. Wong, (2000), "Computation of Value-at-Risk for Nonlinear Portfolios", *Journal of Risk* 3, 37-55.
- Fleming J., C. Kirby et B. Ostdiek, (2001), "The Economic Value of Volatility Timing", *Journal of Finance* 56(1), 329-352.
- Fleming J., C. Kirby et B. Ostdiek, (2002), "The Economic Value of Volatility Timing using Realized Volatility", *Journal of Financial Economics* 67, 473-509.
- Foster D. et D. Nelson, (1996), "Continuous Record Asymptotics for Rolling Sample Variance Estimators", *Econometrica* 64, 139-174.
- Frino A., A. Lepone et E. Jarnećić, (2004), "Bid-Ask Bounce and the Measurement of Price Behaviour around Block Trades on the Australian Stock Exchange", *EFMA 2004 Basel Meetings Paper*, 29 pages.

- Gabaix X., P. Gopikrishnan, V. Plerou et H. Stanley, (2003), "A Theory of Power-law Distributions in Financial Market Fluctuations", *Nature* 423, 267-270.
- Gallo G., (2001), "Modelling the Impact of Overnight Surprises on Intra-daily Volatility", *Australian Economic Papers* 40(4), 567-580.
- Geman H., D. Madan et M. Yor, (2002), "Stochastic Volatility, Jumps and Hidden Time Changes", *Finance and Stochastics* 6, 63-90.
- Gençay R., F. Selçuk et B. Whitcher, (2003), "Systematic Risk and Timescales", *Quantitative Finance* 3, 108-116.
- Gibbons M., S. Ross et J. Shanken, (1989), "A Test of the Efficiency of a Given Portfolio", *Econometrica* 57, 1121-1152.
- Gillemot L., D. Farmer et L. Fabrizio, (2005), "There's more to Volatility than Volume", *Working Paper*, Santa Fe Institute, 25 pages.
- Giot P. et S. Laurent, (2004), "Modelling Daily Value-at-Risk using Realized Volatility and ARCH type Models", *Journal of Empirical Finance* 11, 379-398.
- Glasserman P., P. Heidelberger et P. Shahabuddin, (2002), "Portfolio Value-at-Risk with Heavy-tailed Risk Factors", *Mathematical Finance* 12(3), 239-269.
- Gleason K., C. Lee et I. Mathur, (2000), "An Explanation for the Compass Rose Pattern", *Economics Letters* 68, 127-133.
- Glosten L. et P. Milgrom, (1985), "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders", *Journal of Financial Economics* 14, 71-100.
- Gonzalo J. et J. Olmo, (2004), "Which Extreme Values Are Really Extreme?", *Journal of Financial Econometrics* 2(3), 349-369.
- Goodhart C. et M. O'Hara, (1997), "High Frequency Data in Financial Markets: Issues and Applications", *Journal of Empirical Finance* 4, 73-114.
- Gottlieb G. et A. Kalay, (1985), "Implication of the Discreteness of Observed Stock Prices", *Journal of Finance* 40(1), 135-153.
- Greenwood P. et M. Nikulin, (1996), *A Guide to Chi-squared Testing*, John Wiley & Sons, 304 pages.
- Griffin J. et R. Oomen, (2008), "Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time?", *Econometric Reviews* 27(1), 230-253.

Guillaume D., O. Pictet et M. Dacorogna, (1995), "On the Intra-daily Performance of GARCH Processes", *Working Paper*, Olsen & Associates, 21 pages.

Härdle W., H. Herwartz et V. Spokoiny, (2003), "Time Inhomogeneous Multiple Volatility Modeling", *Journal of Financial Econometrics* 1(1), 55-95.

Harris L., (1986), "Cross-security Tests of the Mixture of Distributions Hypothesis", *Journal of Financial and Quantitative Analysis* 21, 39-46.

Harris L., (1990), "Estimating of Stock Price Variances and Serial Covariances from Discrete Observations", *Journal of Financial and Quantitative Analysis* 25, 291-306.

Harris L., (1991), "Stock Price Clustering and Discreteness", *Review of Financial Studies* 4, 389-415.

Harvey C. et A. Siddique, (2000), "Conditional Skewness in Asset Pricing Tests", *Journal of Finance* 55, 1263-1295.

Hendrics D., (1996), "Evaluation of Value-at-Risk Models using Historical Data", *Economic Policy Review* 2, 39-69.

Ho T. et H. Stoll, (1983), "The Dynamics of Dealer Markets under Competition", *Journal of Finance* 38, 1053-1074.

Holland J., (1998), *Adaptation in Natural and Artificial Systems*, Fifth Edition (First Edition: 1975), M.I.T Press, 211 pages.

Holthausen R., R. Leftwich et D. Mayers, (1987), "The Effect of Large Block Transactions on Security Prices: A Cross-sectional Analysis", *Journal of Financial Economics* 19, 237-268.

Huang R. et H. Stoll, (1994), "Market Microstructure and Stock Return Prediction", *Review of Financial Studies* 7(1), 179-213.

Huang R. et H. Stoll, (2001), "Tick Size, Bid-Ask Spread and Market Structure", *Journal of Financial and Quantitative Analysis* 36(4), 503-522.

Hurlin C. et S. Tokpavi, (2007), "Backtesting Value-at-Risk Accuracy: A Simple New Test", *Journal of Risk* 9(2), 19-37.

Hyvärinen A. et E. Oja, (1997), "A Fast Fixed-point Algorithm for Independent Component Analysis", *Neural Computation* 9, 1483-1492.

Hyvärinen A., (1999), "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis", *IEEE Transactions on Neural Networks* 10(3), 626-634.

- Hyvärinen A., J. Karhunen et E. Oja, (2001), *Independent Component Analysis*, John Wiley & Sons, New York, 476 pages.
- Jagannathan R. et Z. Wang, (1996), "The Conditional CAPM and the Cross-section of Expected Returns", *Journal of Finance* 51, 3-53.
- Jagannathan R. et Z. Wang, (2002), "Empirical Evaluation of Asset-pricing Models: A Comparison of The SDF and Beta Methods", *Journal of Finance* 57, 2337-2367.
- Jondeau E. et M. Rockinger, (2003), "Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements", *Journal of Economic Dynamics and Control* 27, 1699-1737.
- Jondeau E. et M. Rockinger, (2006), "Optimal Portfolio Allocation under Higher Moments", *European Financial Management* 2(1), 29-55.
- Jurczenko E. et B. Maillet, (2006), "The Four-moment Capital Asset Pricing Model: Between Asset Allocation and Asset Pricing", in *Multi-moment Asset Allocation and Pricing Models*, Jurczenko-Maillet Eds, John Wiley & Sons, 113-163.
- Kan R. et C. Zhang, (1999a), "GMM Test of Stochastic Discount Factor Models with useless Factors", *Journal of Financial Economics* 54, 103-127.
- Kan R. et C. Zhang, (1999b), "Two-pass Tests of Asset Pricing Models with useless Factors", *Journal of Finance* 54, 203-235.
- Kan R. et G. Zhou, (1999), "A Critique of the Stochastic Discount Factor Methodology", *Journal of Finance* 54, 1221-1248.
- Karpoff J., (1987), "The Relation between Price Changes and Trading Volume: A Survey", *Journal of Financial and Quantitative Analysis* 22, 109-126.
- Keim D. et A. Madhavan, (1995), "The Anatomy of the Trading Process: Empirical Evidence on the Behaviour of Institutional Traders", *Journal of Financial Economics* 25, 75-98.
- Keim D. et A. Madhavan, (1996), "The Upstairs Market for Large-block Transactions: Analysis and Measurement of Price Effects", *Review of Financial Studies* 9, 1-36.
- Keim D. et A. Madhavan, (1997), "Transaction Costs and Investment Style: An Inter-exchange Analysis of Institutional Equity Trades", *Journal of Financial Economics* 46, 265-292.
- Keith R. et Y. Shen, (2004), "Coastal Flooding and the Multivariate Skew- t Distribution", in *A Journey beyond Normality*, Genton Ed., Chapman & Hall, 243-258.
- Kim S., L. Lockwood et T. McInish, (1996), "Intraday Beta Stability", in *Advances in Quantitative Analysis of Finance and Accounting*, Cheng-Few Eds, Jay Press Inc., 1996, 35-46.

- Koppl R. et C. Nardone, (2001), "The Angular Distribution of Asset Returns in Delay Space", *Discrete Dynamics in Nature and Society* 6, 101-120.
- Koppl R. et S. Tuluca, (2004), "Random Walk Hypothesis Testing and the Compass Rose", *Finance Letters* 2(1), 14-17.
- Kopriva I. et D. Seršić, (2008), "Wavelet Packets Approach to Blind Separation of Statistically Dependent Sources", *Neurocomputing* 71, 1642-1655.
- Krämer W. et R. Runde, (1997), "Chaos and the Compass Rose", *Economic Letters* 54(2), 113-118.
- Kraus A. et R. Litzenberger, (1976), "Skewness Preferences and the Valuation of Risk Assets", *Journal of Finance* 31, 1085-1100.
- Kunitomo N., (1992), "Improving the Parkinson Methods of Estimating Security Price Volatilities", *Journal of Business* 65, 295-302.
- Kupiec P., (1995), "Techniques for Verifying the Accuracy of Risk Measurement Models", *Journal of Derivatives* 2, 173-84.
- Lai Z., Y. Cheung et L. Xu, (1999), "Independent Component Ordering in ICA Analysis of Financial Data", in *Computational Finance 1999*, Abu Mostafa-Lebaron-Lo-Weigend Eds, MIT Press, 201-212.
- Ledoit O. et M. Wolf, (2003), "Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection", *Journal of Empirical Finance* 10, 603-621.
- Ledoit O. et M. Wolf, (2004), "Honey, I Shrunk the Sample Covariance Matrix", *Journal of Portfolio Management* 30, 110-119.
- Lee C., K. Gleason et I. Mathur, (1999), "A Comprehensive Examination of the Compass Rose Pattern in Futures Markets", *Journal of Futures Markets* 19, 541-564.
- Lee C., K. Gleason et I. Mathur, (2005), "The Tick/Volatility Ratio as a Determinant of the Compass Rose Pattern", *European Journal of Finance* 11(2), 93-109.
- Lettau M. et S. Ludvigson, (2001), "Resurrecting the CCAPM: A Cross-sectional Test when Risk Premia are Time-varying", *Journal of Political Economy* 109, 1238-1287.
- Lintner J., (1965), "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets", *Review of Economics and Statistics* 47, 13-37.

- Loretan M., (1997), "Generating Market Risk Scenarios using Principal Component Analysis: Methodological and Practical Considerations", *manuscript*, Federal Reserve Board, 60 pages.
- Lu W. et J. Rajapakse, (2000), "Constrained Independent Component Analysis", in *Advances in Neural Information Processing Systems 13*, Leen-Dietterich-Tresp Eds., Cambridge, MIT Press, 570-576.
- Lu W. et J. Rajapakse, (2005), "Approach and Applications of Constrained ICA", *IEEE Transactions On Neural Networks* 16(1), 203-212.
- Malevergne Y. et D. Sornette, (2006), "Multi-moments Methods for Portfolio Management: Generalized Asset Pricing Model in Homogeneous and Heterogeneous Markets", in *Multi-moment Asset Allocation and Pricing Models*, Jurczenko-Maillet Eds, John Wiley & Sons, 164-193.
- McKenzie M. et A. Frino, (2003), "The Tick/Volatility Ratio as a Determinant of the Compass Rose: Empirical Evidence from Decimalisation on the NYSE", *Accounting & Finance* 43(3), 331-351.
- McNeil A. et R. Frey, (2000), "Estimation of Tail-related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach", *Journal of Empirical Finance* 7, 271-300.
- Meddahi N., (2002), "A Theoretical Comparison between Integrated and Realized Volatilities", *Journal of Applied Econometrics* 17, 475-508.
- Merton R., (1973), "An Intertemporal Capital Asset Pricing Model", *Econometrica* 41, 867-87.
- Mitchell H. et M. McKenzie, (2006), "A Note on the Wang and Wang Measure of the Quality of the Compass Rose", *Journal of Banking & Finance* 30, 3519-3524.
- Moody J. et H. Yang, (1999), "Term Structure of Interactions of Foreign Exchange Rates", in *Computational Finance*, Abu Mostafa-Lebaron-Lo-Weigend Eds, MIT Press, 246-265.
- Morano C., (2007), "Estimating, Filtering and Forecasting Realized Betas", *Journal of Financial Forecasting* 1, 83-111.
- Morel C. et J. Teïletche, (2001), "Existe-t-il des seuils psychologiques sur les marchés boursiers ? Une application au future CAC 40 sur données très fréquentes", *Banque & Marchés* 52, 17-25.
- Nelson D., (1990), "ARCH Models as Diffusion Approximations", *Journal of Econometrics* 45, 7-38.
- Nelson D. et D. Foster, (1994), "Asymptotic Filtering Theory for Univariate ARCH Models", *Econometrica* 62(1), 1-41.

Nielsen M. et P. Frederiksen, (2008), "Finite Sample Accuracy and Choice of Sampling Frequency in Integrated Volatility Estimation", *Journal of Empirical Finance* 15, 265-286.

Oomen R., (2002), "Modelling Realized Variance when Returns are Serially Correlated", *Working Paper*, Warwick Business School, 33 pages.

Oomen R., (2002), "Statistical Models for High Frequency Security Prices", *Working Paper*, Warwick Business School, 45 pages.

Parks R., (1967), "Efficient Estimation of a System of Regression Equations when Disturbances are both Serially and Contemporaneously Correlated", *Journal of the American Statistical Association* 62, 500-509.

Protter P., (2004), *Stochastic Integration and Differential Equations*, Second Edition, Springer-Verlag, New York, 444 pages.

Rodgers J. et W. Nicewander, (1988), "Thirteen Ways to Look at Correlation Coefficient", *The American Statistician* 42(1), 59-66.

Rojas S. et J. Moody, (2001), "Cross-sectional Analysis of the Returns of iShares MSCI Index Funds using Independent Component Analysis", *Working Paper*, Oregon Graduate Institute of Science and Technology, 48 pages.

Ross S., (1976), "The Arbitrage Theory of the Capital Asset Pricing", *Journal of Economic Theory* 13, 341-360.

Ruppert D., (1987), "What is Kurtosis? An Influence Approach", *The American Statistician* 41(1), 1-5.

Rydberg T. et N. Shephard, (2003), "Dynamics of Trade-by-trade Price Movements: Decomposition and Models", *Journal of Financial Econometrics* 1(1), 2-25.

Scholes M. et J. Williams, (1977), "Estimating Betas from Non-synchronous Data", *Journal of Financial Economics* 5, 309-327.

Shanken J., (1992), "On the Estimation of Beta Pricing Models", *Review of Financial Studies* 5, 1-33.

Sharpe W., (1964), "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", *Journal of Finance* 19, 425-442.

Smith D., (2007), "Conditional Coskewness and Asset Pricing", *Journal of Empirical Finance* 14, 91-119.

- Szipiro G., (1998), "Tick Size, the Compass Rose and Market Nanostructure", *Journal of Banking and Finance* 22(12), 1559-1569.
- Tsai K., (2004), "Risk Management *via* Value-at-Risk", *Working Paper*, Aventis Pharmaceuticals, 10 pages.
- Valle A., (2004), "The Skew-Normal Distribution", in *A Journey beyond Normality*, Genton Ed., Chapman & Hall, 3-24.
- Voit J., (2003), *The Statistical Mechanics of Financial Markets*, Second Edition, Springer, 287 pages.
- Vorlow C., (2004), "Stock Price Clustering and Discreteness: The Compass Rose and Complex Dynamics", *Working Paper*, University of Durham, 17 pages.
- Wang E., R. Hudson et K. Keasey, (2000), "Tick Size and the Compass Rose: Further Insights", *Economics Letters* 68, 119-125.
- Wang H. et C. Wang, (2002), "Visibility of the Compass Rose in Financial Asset Returns: A Quantitative Study", *Journal of Banking & Finance* 26(6), 1099-1111.
- Woehrmann P., (2007), "Discrete Stock Prices and Predictability", *Working Paper* 245, National Centre of Competence in Research, 20 pages.
- Wood R., T. McInish et J. Ord, (1985), "An Investigation of Transactions Data for NYSE Stocks", *Journal of Finance* 40(3), 723-741.
- Zellner A., (1962), "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", *Journal of the American Statistical Association* 57, 348-368.
- Zhang B., (2003), "Time-varying Volatilities, CAPM Betas, and Factor Loadings: A High Frequency Data Perspective", *Unpublished PhD thesis*, Duke University, 120 pages.
- Zhang K. et L. Chan, (2006), "An Adaptive Method for Subband Decomposition ICA", *Neural Computation* 18, 191-223.
- Zumbach G., F. Corsi et A. Trapletti, (2002), "Efficient Estimation of Volatility Using High Frequency Data", *Working Paper*, Olsen & Associates, 22 pages.